

Signal Processing and Classification Approaches for Brain-computer Interface

Tarik Al-ani^{1,2} and Dalila Trad^{1,3}

¹LISV-UVSQ, 10-12 Av de l'Europe, 78140 Velizy

²Department of Informatics, ESIEE-Paris, Cité Descartes-BP 99 93162 Noisy-Le-Grand

³UTIC-ESSTT, University of TUNIS 5, avenue Taha Hussein,
B.P. 56 Bab Menara 1008- Tunis

^{1,2}France

³Tunisia

1. Introduction

Research on *brain-computer interface* (BCI) systems began in the 1970s at the University of California Los Angeles (UCLA) (Vidal, 1973; 1977). The author gave in his papers the expression "Brain Computer Interface" which is the term currently used in literature.

A BCI system is a direct communication pathway between a brain and an external artificial device. BCI systems were aimed at assisting, augmenting or repairing *human cognitive* or *sensory-motor functions*.

The BCI systems (BCIs) allow control of an artificial device based on the features extracted from voluntary electric, magnetic, or other physical manifestations of brain activity collected from epi- or subdurally from the cortex or from the scalp or in invasive electrophysiological manner, i.e. brain signals recorded intracortically with single electrode or multi-electrode arrays (Dornhege et al., 2007). There is a variety of non-invasive techniques for measuring brain activity. These non-invasive techniques include, the *electroencephalography* (EEG), *magnetoencephalography* (MEG), *positron emission tomography* (PET), *functional magnetic resonance imaging* (fMRI), and *optical imaging*. However, for technical, time resolution, real-time, and price constraints, only EEG monitoring and related techniques are employed in the BCI community. For more details refer to (Wolpaw et al., 2002; Mason et al., 2007; Dobkin, 2007). The neuronal electrical activity contain a broad band frequency, so the monitored brain signals are filtered and denoised to extract the relevant information (see section 3) and finally this information is decoded (see section 6) and commuted into device commands by *synchronous control* or more efficiently by *self-paced* or *asynchronous control* in order to detect whether a user is intending something or not (see chapter 7 in (Dornhege et al., 2007) for details), Fig. 1. For some specific BCI tasks, raw brain signal serves as stimulus as well as a control interface feedback.

The direct BCIs can be seen as a new means of communication that may be used to allow tetraplegic or individuals with severe motor or neuromuscular diseases (e.g. Amyotrophic lateral sclerosis (ALS), brainstem stroke, brain or spinal cord injury, cerebral palsy, muscular

Source: Intelligent and Biosensors, Book edited by: Vernon S. Somerset,
ISBN 978-953-7619-58-9, pp. 386, January 2010, INTECH, Croatia, downloaded from SCIYO.COM

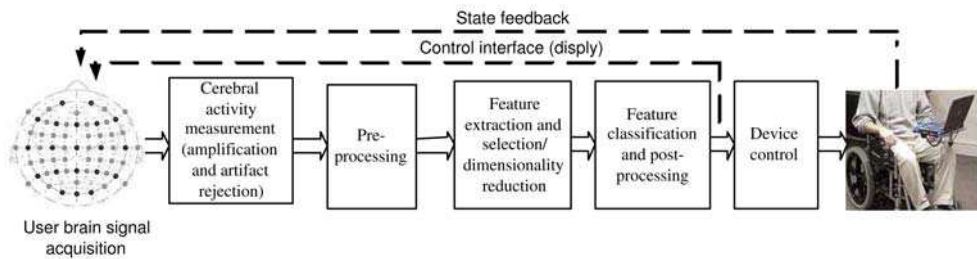


Fig. 1. Basic BCI layout.

dystrophies, multiple sclerosis) to have effective control over artificial devices or external environment in order to increase or improve their communication qualities or their independence. Recent studies have demonstrated correlations between EEG signals and actual or imagined movements and between EEG signals and mental tasks (Keirn & Aunon, 1990; Lang et al., 1996; Pfurtscheller et al., 1997; Anderson et al., 1998; Altenmüller & Gerloff, 1999; McFarland et al., 2000; Wessberg et al., 2000; Pfurtscheller et al., 2000b; Nicoletis, 2001; Pfurtscheller et al., 2003). The BCIs can be used also in therapeutic applications by neurofeedback for rehabilitation or functional recovery (Birbaumer & Cohen, 2007; Dobkin, 2007; Birbaumer et al., 1999; Dornhege et al., 2007).

The BCI is a communication system that does not require any peripheral muscular activity. It has been shown by (Pfurtscheller & Aranibar, 1977; Pfurtscheller, 1999c; Neuper & Pfurtscheller, 1999a) that the imagination of either a left or right hand movement results in an amplitude attenuation (*event-related desynchronization* (ERD) of μ (8-13Hz) and central β (13-30Hz) rhythms at the contra-lateral sensorimotor representation area and, in an amplitude increase (*event-related synchronization* (ERS) within the γ band (30-40Hz) at the ipsi-lateral hemisphere. The event related (de)synchronisation (ERD, ERS) (Pfurtscheller et al., 1999a), see Fig. 2 and Fig. 3.

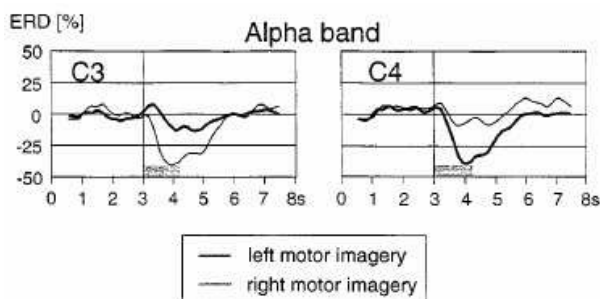


Fig. 2. Grand average ERD curves recorded during motor imagery from the left (C3) and right sensorimotor cortex (C4) (the electrodes C3 and C4 are placed according to the International 10-20 system). The ERD time courses were calculated for the selected bands in the alpha range for 16 subjects. Positive and negative deflections, with respect to baseline (second 0.5 to 2.5), represent a band power increase (ERD) and decrease (ERD), respectively. The gray bar indicates the time period of cue presentation (i.e. the imagination starts at second 3). Figure from (Pfurtscheller et al., 2000a) which is modified from (Neuper & Pfurtscheller, 1999a).

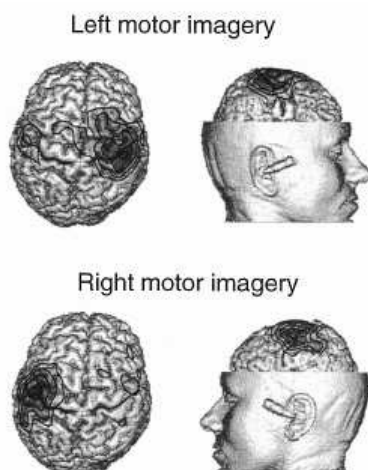


Fig. 3. ERD maps for a single subject calculated for the cortical surface of a realistic head model. Figure from (Pfurtscheller et al., 2000a) which is modified from (Neuper & Pfurtscheller, 1999a).

The direct *BCIs* can also be seen as a new means to extend communication for healthy subjects in many fields such as multimedia communication, control of robots, virtual reality and video games (Thomas, 1977; Friedman et al., 2004; Bell et al., 2008; Lécuyer et al., 2008). There are in general two types of *BCI* systems: *endogenous tasks* and *exogenous tasks* based systems (Dornhege et al., 2007).

The *endogenous tasks BCI* systems, which are based on spontaneous activity, use brain signals that do not depend on external stimuli and that can be influenced by concentrating on a specific mental task. In order to obtain an efficient task recognition system, several concentration trials of human are, in general, realized. The concentration constraint is a very tiring mental task especially for disabled subjects who might have difficulties in acquiring voluntary control over their brain activity and it must be reduced in order to obtain an efficient task recognition system.

The *exogenous tasks BCI* systems, which are based on evoked activity, use brain signals that do depend on external stimuli. Particularly interesting are systems based either on the *P300* or on *SSVEPs* (see section 2). Advantages of these potentials are that they are relatively well understood from a neurophysiologic point of view and that they can be evoked robustly across different subjects. Moreover, feedback training is not necessary in these systems, as these potentials appear "automatically" whenever subjects concentrate onto one out of several stimuli presented in random order (Hoffman et al., 2008). Note that the material presented in this chapter is strongly biased towards *sensorimotor* (Changes in brain rhythms (μ , β , and γ)) and *P300* electrophysiological activities using *EEG* records.

In order to improve the performance of the *BCI* system design, it is necessary to use a good method of *signal processing* to allow easier extraction of physiological characteristics and also to use a good *classifier* adapted to the specificities of the *BCI* system. This chapter presents a compact guide to different *signal processing* techniques that have received more attention in *BCIs*. We introduce then some selected feature extraction and *classification* approaches in the context of *BCI* systems. A more exhaustive and excellent surveys on *signal processing* and

classification algorithms may be found in the papers (Bashashati et al., 2007; Lotte et al., 2007). Then this chapter describes the application of two *classification* approaches, *hidden Markov models* (HMMs) and *support vector machines* (SVM), in the context of *exogenous tasks* BCI systems based on *P300 evoked potential*. The chapter ends with a global conclusions and perspectives.

The methods presented in sections 3.3, 4, 5 and 6 are based on the statistical results given in the comprehensive survey of 96 BCI designs using electrical signal recordings published prior to January 2006 by (Bashashati et al., 2007). Among these methods, we give here only a brief descriptions of the most applied methods. They are introduced here without referencing all the published papers for the 96 BCI designs. The reader may refer to the paper (Bashashati et al., 2007) to find a rich bibliographical work. However, we give only the original references corresponding to each proposed method.

2. Electrophysiological control activities in BCIs

Current BCI systems fall into seven main categories, based on the neuromechanisms and recording technology they use to generate control signals (Bashashati et al., 2007). The following list give a short descriptions of these electrophysiological activities used in BCI designs. This list is borrowed and adapted (with the authorization of authors) from the paper (Bashashati et al., 2007). We omitted the references of the different approaches given in this list. Many of these references are given in (Bashashati et al., 2007).

- **Sensorimotor activity** BCI designs that use sensorimotor activity as the neural source of control can be divided into three sub-categories:
 - **Changes in brain rhythms (μ , β , and γ)**
 μ rhythms in the range of 8-12 Hz and β rhythms in the range of 13-30 Hz both originate in the sensorimotor cortex and are displayed when a person is not engaged in processing sensorimotor inputs or in producing motor outputs. They are mostly prominent in frontal and parietal locations. A voluntary movement results in a circumscribed desynchronization in the μ and lower β bands. This desynchronization is called *event-related desynchronization* (ERD) (Pfurtscheller & Aranibar, 1977; Pfurtscheller, 1999c; Neuper & Pfurtscheller, 1999a) and begins in the contralateral rolandic region about 2 s prior to the onset of a movement and becomes bilaterally symmetrical immediately before execution of movement. After a voluntary movement, the power in the brain rhythms increases. This phenomenon, called *event-related synchronization* (ERS), is dominant over the contralateral sensorimotor area and reaches a maximum around 600 ms after movement offset. γ rhythm is a high-frequency rhythm in the EEG. Upon the occurrence of a movement, the amplitude of γ rhythm in the range of 30-40 Hz increases. Gamma γ are usually more prominent in the primary sensory area.
 - **Movement-related potentials (MRPs)**
 MRPs are low-frequency potentials that start about 1-1.5 s before a movement. They have bilateral distribution and present maximum amplitude at the vertex. Close to the movement, they become contralaterally preponderant.
 - **Other sensorimotor activities**
 The sensorimotor activities that do not belong to any of the preceding categories are categorized as other sensorimotor activities. These activities are usually not restricted to a particular frequency band or scalp location and usually cover

different frequency ranges. An example would be features extracted from an *EEG* signal filtered to frequencies below 30 Hz. Such a range covers different *eventrelated potentials* (ERPs) but no specific neuromechanism is used.

- **Slow cortical potentials (SCPs)**

Slow cortical potentials (SCPs) are slow voltage shifts in the *EEG* occurring in the frequency range 1-2 Hz. Negative SCPs correspond to a general decrease in cortical excitability. Positive SCPs correspond to a general increase in cortical excitability. Through feedback training subjects can learn to voluntarily control their SCPs (Birbaumer et al., 1999; 2000; Hinterberger et al., 2003; 2004; Bostanov, 2004).

- **P300 Evoked potential**

Infrequent or particularly significant auditory, visual, or somatosensory stimuli, when interspersed with frequent or routine stimuli, typically evoke in the *EEG* over the parietal cortex a positive peak at about 300 ms after the stimulus is received. This peak is called *P300*.

- **Visual evoked potentials (VEPs)**

VEPs are small changes in the ongoing brain signal. They are generated in response to a visual stimulus such as flashing lights and their properties depend on the type of the visual stimulus. These potentials are more prominent in the occipital area. If a visual stimulus is presented repetitively at a rate of 5-6 Hz or greater, a continuous oscillatory electrical response is elicited in the visual pathways. Such a response is termed steady-state visual evoked potentials (SSVEP). The distinction between VEP and SSVEP depends on the repetition rate of the stimulation.

- **Response to mental tasks**

BCI systems based on non-movement mental tasks assume that different mental tasks (e.g., solving a multiplication problem, imagining a 3D object, and mental counting) lead to distinct, task-specific distributions of *EEG* frequency patterns over the scalp.

- **Activity of neural cells (ANC)**

It has been shown that the firing rates of neurons in the motor cortex are increased when movements are executed in the preferred direction of neurons. Once the movements are away from the preferred direction of neurons, the firing rate is decreased.

- **Multiple neuromechanisms (MNs)**

BCI systems based on multiple neuromechanisms use a combination of two or more of the above mentioned neuromechanisms.

3. Signal pre-processing methods in BCIs

To extract *features* (see section 4), it is necessary to pre-process first the data. Three steps are necessary to achieve this goal: *Referencing*, *Temporal filtering* and *signal enhancement*.

3.1 Referencing

(Hagemann et al., 2001) have stated that the differences between results of different studies are partly due to the differences in referencing. In the case of *EEG* recordings from the cortex or from the scalp, these recordings are obtained using, in general, different electrodes on different positions. Since the brain activity voltage measured by a given electrode is a relative measure, the measurement may be compared to another reference brain voltage situated on another site. This results in a combination of brain activity at the given electrode, brain activity at the reference site and noise. Because of this, the reference site should be

chosen such that the brain activity at that site is almost zero. Typically, the nose, mastoids and earlobes are used (Dien, 1998). In general, there are three referencing methods

- **Common reference**

The common reference technique is widely used in *BCIs*. This method uses one common reference for all electrodes. In general, the site of this reference is situated at large distance from all electrodes. The activity at the reference site influences all measurements equally, and differences between electrode measurements still contain all information needed.

- **Average reference**

The average reference subtracts the average of the activity at all electrodes from the measurements. This method is based on the principle that the activity at the whole head at every moment sums up to zero. Therefore, the average of all activity represents an estimate of the activity at the reference site. Subtracting this average produces in principle a dereferenced solution. However, the relatively low density of the electrodes and the fact that the lower part of the head is not taken into account, bring some practical problems along (Dien, 1998).

- **Current source density (CSD)**

The current source density (*CSD*) is used in many *BCIs*. It is "the rate of change of current flowing into and through the scalp" (Weber, 2001). This quantity can be derived from *EEG* data, and it may be interpreted as the potential difference between an electrode and a weighted average of their surrounding electrodes. The *CSD* can be estimated by computing the *laplacian*. The *laplacian* computes the sum of the differences between an electrode and its neighbours. A problem with this estimation is that it is actually only valid when the electrodes are in a two dimensional plane and equally distant.

3.2 Temporal filtering in *BCIs*

The brain signals are naturally contaminated by many internal and external noises. They can be removed using simple filters. The relevant information in *BCIs* is found in the frequencies below 30Hz. Therefore, all noise with higher frequencies (e.g. noise from the electrical net has a fixed frequency of 50Hz or 60 Hz) can be removed using *FIR low pass filter*. Specific frequency bands may also be selected using *FIR bandpass filters*.

3.3 Signal enhancement methods in *BCI* designs

The choice of a suitable enhancement technique is dependent on several factors such as the recording technology, number of electrodes, and neuromechanism of the *BCI* (Bashashati et al., 2007). Among seventeen pre-processing methods given by (Bashashati et al., 2007), we describe here briefly only six methods which are the most applied in *BCI* designs:

- **Spatial filters - Referencing methods**

The proper selection of a spatial filter for any *BCI* is determined by the location and extent of the selected brain control signal and of the various sources of *EEG* or *non-EEG* noise.

- **Common average referencing (CAR)**

Common-average or "reference-free" recording has been suggested as a solution to the problem of the reference electrode (Offner, 1950; lehmann & Skrandies, 1984; Stanny, 1989). Common-average referencing involves recording in bipolar fashion from a number of electrodes, all referred to a single site. One then calculates the

grand mean *EEG* waveform, by averaging across electrodes, and subtracts the result pointwise from the *EEG* recorded at each electrode. Activity recorded by the reference electrode is theoretically of equal magnitude in the mean and individual electrode waveforms. Consequently, the effect of the reference electrode should be eliminated from each recording electrode's output when the common-average waveform is subtracted (Stanny, 1989).

- **Surface Laplacian (SL)**

The *SL* is defined as the 2nd order spatial derivative of the surface potential. Due to its intrinsic spatial high-pass filtering characteristics, the *SL* can reduce the volume conduction effect by enhancing the high-frequency spatial components, therefore can achieve higher spatial resolution than surface potentials.

- **Principal component analysis (PCA)**

The *PCA* (Pearson, 1901) is a linear mapping that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Depending on the field of application, it is also named the discrete *Karhunen-Loève transform (KLT)*, the *Hotelling transform* or *proper orthogonal decomposition (POD)*. The *PCA* reveals the internal structure of the data in a way which best explains the variance in the data. If a multivariate dataset is visualised as a set of coordinates in a high-dimensional data space (1 axis per variable), *ICA* supplies the user with a lower-dimensional representation.

- **Independent component analysis (ICA)**

The more important *artefacts* in *BCIs* are generated by muscles and eyes blink (Gupta & Singh, 1996). Classical automatic methods for removing such *artefacts* can be classified into *rejection* methods and *subtraction* methods.

- **Rejection** methods consist of discarding contaminated *EEG*, based on either automatic or visual detection can be used in the *BCI* applications framework. Their success crucially depends on the quality of the detection.
- **Subtraction** methods are based on the assumption that the contaminated *EEG* is a linear combination of an original *EEG* and other independent *artefact* signals generated by the muscles and eyes blink. The original *EEG* is hence recovered by either subtracting separately recorded *artefact*-related signals from the measured *EEG*, using appropriate weights or by applying recent approaches for *artefacts rejection*: such as *independent component analysis (ICA)* (Common, 1994; Hyvärinen & Oja, 2000), *peak elimination* (Nakamura et al., 1996), *neural network* (Urszula et al., 1999) and *fixed bandpass FIR filter* based approach (Gupta & Singh, 1996).

The *ICA* (Common, 1994; Hyvärinen & Oja, 2000) is the more used technique. It is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals. It is a special case of *blind source separation (BSS)*. *ICA* is particularly efficient when the *EEG* and the *artefacts* have comparable amplitudes. For more details about their advantages, their limitations and their applications for the removal of eyes activity *artefacts*, refer to (Jung et al., 1998; 2000).

- **Common spatial patterns (CSP)**

The *CSP* (Koles, 1991; Müller-Gerking et al., 1999) is a technique used to find the common projection matrix that decomposes the different classes of single trial *EEG*

datasets, and more specifically to find spatial structures of event-related (de)synchronization (ERD/ERS) in a EEG context. Such matrix maximizes the differences between the classes. (Guger et al., 2000) demonstrated the efficiency of the CSP method for real-time EEG analysis and concluded that only parameters that must be adjusted for the CSP are the time segment for the calculation of the CSP and, during on-line processing, the time window for the calculation of the variances. But the selection of these parameters is not very crucial. An advantage of the CSP method is that it does not require a priori selection of subject-specific frequency bands, as necessary for bandpower or frequency estimation methods (Pfurtscheller et al., 1996; McFarland et al., 1997b).

- The CSP method is very sensitive to artefacts. A single trial containing, for example, a movement artifact can cause severe changes in the CSP (Müller-Gerking et al., 1999). The reason is the sample covariance (nonrobust estimate), which is used to estimate the covariance for the calculation of the spatial filters. However, during on-line operation of the BCI, the spatial filters perform a weighted spatial averaging of the EEG, and this reduces the influence of artefacts (Guger et al., 2000).
- in some applications, many electrodes are needed, (e.g. more than 18 (Ramoser et al., 2000)), which necessitates costly hardware.
- since the CSP method detects spatial patterns in the EEG, any change in the electrode positions may render the improvements in the *classification* accuracy gained by this method useless. Therefore, this method requires almost identical electrode positions for all trials and sessions which may be difficult to accomplish (Ramoser et al., 2000). (Guger et al., 2000) recommended not to apply the electrodes anew after setting up a new CSP for the following feedback sessions. For long-term implications to analyze the EEG in real time, EEG data of several sessions can be used for the calculation of the CSP. This allows the generation of a more robust filter in order to overcome the mentioned problems.
- **Common spatial subspace decomposition (CSSD)**
The CSSD can extract signal components specific to one condition from multiple MEG/EEG data sets of multiple task conditions. Signal matrices or covariance matrices are decomposed using spatial factors common to multiple conditions. The spatial factors and corresponding spatial filters are then dissociated into specific and common parts, according to the common spatial subspace which exists among the data sets. Finally, the specific signal components are extracted using the corresponding spatial filters and spatial factors. (Wang et al., 1999).
- **Frequency normalization (Freq-Norm)**
(Bashashati et al., 2005).
- Other methods are given by (Bashashati et al., 2007).

The study of (Bashashati et al., 2007) showed that

- signal pre-processing algorithms have been used for EEG-based BCIs and the ANC-based BCIs, but no signal enhancement algorithms have been applied on *electrocorticogram* (ECoG)-based BCIs. Only PCA has been used in both groups, and
- *spatial filtering* including *referencing* (CAR and SL) methods and CSP are among the most used techniques that have become increasingly popular in EEG-based BCIs.

Fig.4 shows the statistical results of the study realised by (Bashashati et al., 2007) concerning *pre-processing* methods in BCI designs. (Bashashati et al., 2007) concluded that 96 BCI designs

that employ signal enhancement techniques before extracting the features from the signal, 32% use surface Laplacian (*SL*), 22% use either principal component analysis (*PCA*) or independent component analysis (*ICA*), 14% use common spatial patterns (*CSP*) and 11% use common average referencing (*CAR*) techniques.

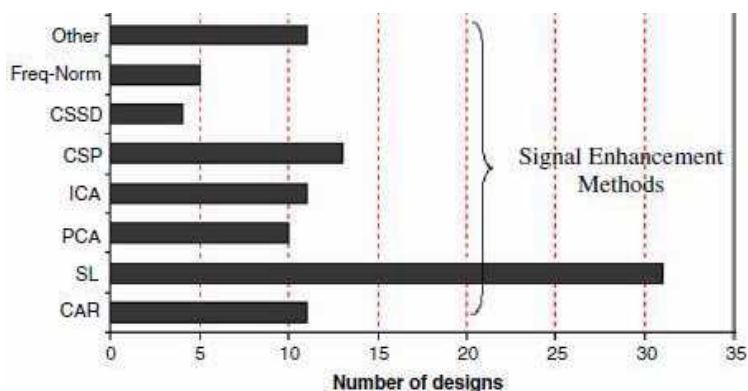


Fig. 4. Signal enhancement methods in *BCI* designs. Figure modified from (Bashashati et al., 2007) with authorisation.

In the following, we give a brief description of the two most used methods: *Spatial filters* and *Common spatial patterns*.

3.3.1 Spatial filters: (SL) and (CAR)

(McFarland et al., 1997a) showed that the variability of the *EEG* or *non-EEG* noise sources within the different *BCI* designs and even within individuals make difficult the application of the spatial filters. For *BCIs* that use the μ and β rhythms, the *SL* and *CAR* methods are superior to the ear reference method. However, it was shown that the reference method (*CAR*, *bipolar*, *large Laplacian*, *small Laplacian*, and *referenced to the ear* (McFarland et al., 1997a)) had minor influence on the *classification* accuracy (Ramoser et al., 2000). Fast and continuous feedback can also enhance the performance of the system (Guger et al., 2001; Neuper et al., 1999b). In the following, we introduce only the principles of the *CSP* given in (Guger et al., 2000).

3.3.2 Common spatial patterns (CSP)

As described by (Guger et al., 2000), the *CSP* method uses the covariance to design *common spatial patterns* and is based on the simultaneous diagonalisation of two covariance matrices (Fukunaga, 1972). The decomposition (or filtering) of the *EEG* leads to new time series, which are optimal for the discrimination of two populations (or classes). The patterns are designed such that the signal resulting from the *EEG* filtering with the *CSP* has maximum variance for population and minimum variance for the second population and vice versa. In this way, the difference between the first and second populations is maximized, and the only information contained in these patterns is where the variance of the *EEG* varies most when comparing two conditions.

Given N channels of *EEG* for each trial \mathbf{X} of population 1 and population 2, the *CSP* method gives an $N \times N$ projection matrix according to (Koles, 1991; Müller-Gerking et al., 1999; Ramoser et al., 2000; Guger et al., 2000). This matrix is a set of subject-specific spatial patterns, which reflect the specific activation of cortical areas during hand movement imagination. With the projection matrix \mathbf{W} , the decomposition of a trial \mathbf{X} is described by

$$\mathbf{Z} = \mathbf{W}\mathbf{X}. \quad (1)$$

This mapping projects the variance of \mathbf{X} onto the rows of \mathbf{Z} and results in new time series. The columns of \mathbf{W}^{-1} are a set of *CSPs* and can be considered as time-invariant *EEG* source distributions. After interpolation, the patterns can be displayed as topographical maps. By construction, the variance for population 1 is largest in the first row of \mathbf{Z} and decreases with the increasing number of the subsequent rows. The opposite is the case for a trial with population 2.

4. Feature extraction methods in BCI designs

This section describes briefly the common *BCI* features extraction methods. Concerning the design of a *BCI* system, some critical properties of these features must be considered (Lotte et al., 2007):

- noise and outliers: the brain signals (e.g. *EEGs*) have a poor signal-to-noise ratio;
- high dimensionality: in *BCI* systems, feature vectors are often of high dimensionality. Several features are generally extracted from several channels and from several time segments before being concatenated into a single feature vector;
- time information: *BCI* features should contain time information as brain activity patterns are generally related to specific time variations of *EEG*;
- the brain signals are non-stationary in nature;
- the brain signals are non-linear in nature;
- non sufficient training sets: training process is time consuming and demanding for the subjects.

There are many methods used in *BCI*, depending of the type of the *BCI* systems. In the following we describe some main and specific methods. More exhaustive details are given by (Bashashati et al., 2007). The feature extraction methods described here are: *Band powers* (BP), *Cross-correlation between EEG band powers*, *frequency representation* (FR), *time-frequency representation* (TFR), *Hjorth parameters*, *parametric modelling*, *inverse model* and specific techniques used for *P300* and *VEP* such as *Peak picking* (PP) and *Slow cortical potentials calculation* (SCPs).

4.1 Band powers (BP)

The features may be extracted from the *EEG* signals by estimating the power distribution of the *EEG* in predefined frequency bands. (Pfurtscheller et al., 1997) used the *band powers* (BP) and demonstrated that for each subject, different frequency components in the α and β band were found which provided best discrimination between left and right hand movement imagination. These frequency bands varied between 9 and 14 Hz and between 18 and 26 Hz.

4.2 Cross-correlation between EEG band powers

In the case of *EEG* measurements the cross-correlation coefficients between the *EEG* activity may be calculated to obtain some information from comparing different locations and

different frequency bands (Farwell & Donchin, 1988; Musha et al., 1997; Bayliss & Ballard, 1999; 2000a;b; Wang et al., 2004a;b).

4.3 Frequency representation (FR)

Frequency representation (FR) features have been widely used in signal processing because of their ease of application, computational speed and direct interpretation of the results (Wolpaw et al., 2000; Blankertz et al., 2006). Specifically, about one-third of *BCI* designs have used *power-spectral density (PSD)* features (Bashashati et al., 2007).

4.4 Time-frequency representation (TFR)

Due to the non-linearity and non-stationarity nature of the *EEG* signal, the classical methods based on *Fourier transform (FT)* are, in general, not efficient for feature extraction because the obtained features do not provide any time domain information, i.e. these features do not analyze the time-varying spectral content of the signals.

Time-frequency methods decompose the *EEGs* into a series of frequency bands, and the instantaneous power is represented by the envelop of oscillatory activity, which forms the spatial patterns for a given electrode montage at a time-frequency grid (Millán & Mouriño, 2003; Wang et al., 2004a).

Wavelet-based feature extraction algorithms (Qin & He, 2005; Xu & Song, 2008; Haibin et al., 2008) necessitate the choice of a particular wavelet called *mother wavelet* in order to extract useful information. This choice of an appropriate *mother wavelet* may be simplified by the prior knowledge of the physiological activity in the brain.

(Huang et al., 1998) proposed a more fairly recent technique called the *Empirical Mode Decomposition (EMD)* was proposed for nonlinear and non-stationary time series data. The (*EMD*) is a data driven approach (i.e. one does not need to define a *mother wavelet* beforehand) that can be used to decompose adaptively a signal into a finite well-defined high frequency and low frequency components, which are known as intrinsic mode functions (*IMFs*) or modes. They consider signals at their local oscillations, but they are not necessarily considered in the sense of Fourier harmonics. Their extraction is non-linear, but their recombination for exact reconstruction of the signal is linear. We think that this approach might be useful in *BCI* design.

4.5 Hjorth parameters

(Hjorth, 1970) has described three parameters that characterize the temporal dynamics of *EEG* signal, ($X(t)$), of length N in terms of *amplitude*, *time scale* and *complexity*. The parameters are measured in the time domain, as opposed to the other features, which are measured in the frequency domain. It has been shown that these parameters are capable of discriminating between different mental states (Vourkas et al., 2000). The parameters are:

- *Activity*: a measure of the mean power of the signal (variance of X ($VAR(X)$)). It is measured as the standard deviation.

$$Activity(X(t)) = VAR(X(t)) = \frac{\sum_{n=1}^N (X(n) - \bar{X})^2}{N}, \quad (2)$$

where \bar{X} denotes the mean of X .

- *Mobility*: represents the mean frequency in the signal. The mobility can be computed as the ratio of the standard deviation of the slope and the standard deviation of the amplitude.

$$Mobility(X(t)) = \sqrt{\frac{Activity(\frac{dX(t)}{dt})}{Activity(X(t))}} = \sqrt{\frac{VAR(\frac{dX(t)}{dt})}{VAR(X(t))}}. \quad (3)$$

- *Complexity*: tries to capture the deviation from the sine wave. It is expressed as the number of standard slopes actually seen in the signal during the average time required for one standard amplitude, as given by the mobility.

$$Complexity(X(t)) = \frac{Mobility(\frac{dX(t)}{dt})}{Mobility(X(t))}. \quad (4)$$

These parameters are mainly used as features for the *classification* of motor imagery (Obermeier et al., 2001; Boostani & Moradi, 2004; Lee & Choi, 2003; Pfurtscheller & Neuper, 2001).

4.6 Parametric modelling

In *statistics*, a *parametric model* or *parametric family* or *finite-dimensional model* refers to a family of distributions which can be described using a finite number of parameters. These parameters are usually collected together to form a single k -dimensional parameter vector $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$. In *system theory*, *parametric model* assume that the time series under analysis to be the output of a given linear mathematical model. They require an a priori choice of the structure and order of the signal generation mechanism model.

Among the more used parametric modelling in *BCIs* are the *autoregressive parameters* (AR) and their variants such as *multivariate AR parameters* (MVAR), *AR parameters with exogenous input* (ARMAX) and *Adaptive AR parameters* (AAR) (Anderson & Sijercic, 1996; Schlogl et al., 1997; Anderson et al., 1998; Roberts & Penny, 2000; Burke et al., 2005; Vidaurre et al., 2007).

AR methods assume that a signal $X(t)$, measured at time t , can be modeled as a weighted sum of the values of this signal at previous time steps, to which we can add a noise term E_t (generally a *Gaussian white noise*):

$$X(t) = a_1X(t-1) + a_2X(t-2) + \dots + a_pX(t-p) + E_t, \quad (5)$$

where the weights a_i are the *autoregressive parameters* which are generally used as features for *BCI*. AAR assume that the weights a_i can vary over time. It seems that (AAR) parameters would give better results than (AR) parameters for motor imagery *classification* (Schlögl et al., 1997; Pfurtscheller et al., 1998), whereas they would give worse results for the *classification* of cognitive tasks such as mental computations, mental rotation of a geometric figure, etc. (Huan & Palaniappan, 2004a; Huan & Palaniappan, 2004b). It should be noted that it is possible to derive a frequential information from the a_i coefficients (McFarland & Wolpaw, 2005).

4.7 Inverse model

Inverse models have shown to be promising feature extraction algorithms (Qin et al., 2004; Grave et al., 2005; Wentrup et al., 2005; Congedo et al., 2006). Such models are able to compute the activity in the whole brain volume, only using *EEG* and a head model that generally represents the brain as a set of *volume elements* (*voxels*). The activity thus calculated in some relevant brain regions or *voxels* may be used as efficient features for *BCI* systems.

4.8 specific techniques

4.8.1 Peak picking (PP)

Peak picking (PP) method detects a specific pattern based on its peak value in a region associated with a specific cognitive component. It is used specifically for the evoked potential *P300* (or *P3*)-based BCI system (Meinicke et al., 2003; Garrett et al., 2003; Bayliss et al., 2004; Bayliss & Inverso, 2005; Salimi Khorshidi et al., 2007; Hoffman et al., 2008), Fig.5.

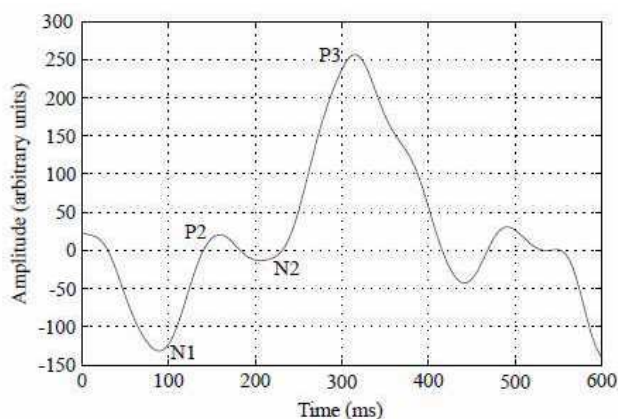


Fig. 5. Typical P300 wave. From (Hoffman et al., 2008).

PP is a simple algorithm to recognize a *P300* component using the difference between the minimum and maximum amplitude in a trial. A trial with a prototypical evoked potential *P300* component contains a large peak from 300-400 ms and *PP* recognizes the *P300* signal when the amplitude difference is greater than or equal to a specified voltage difference between the minimum, $\min(x)$, and maximum, $\max(x)$, voltage points within a specified time window, where x is a vector which represents the data for a single *P300* response. For recognition, the time window with the best results may be between three and six hundred milliseconds citepbayliss04.

4.8.2 Slow cortical potentials (SCPs) calculation methods

The *SCPs* amplitudes are extracted on-line from the regular electroencephalogram, filtered, corrected for eye movement artefacts and fed back to the patient with visual, auditory or tactile feedback (Birbaumer et al., 1999; 2000; Hinterberger et al., 2004). The *TFR* methods are also used to extract the features of the *SCPs* (Hinterberger et al., 2003; Bostanov, 2004). Fig.6 shows the statistical results of the study realised by (Bashashati et al., 2007) concerning *feature extraction* methods in BCI designs. (Bashashati et al., 2007) concluded that 41% of the BCIs are based on the sensorimotor activity use *PSD* features, 16% rely on parametric modelling of the data, 13% use *TFR* methods and 6% do not employ any feature extraction methods. 74% of the *SCP*-based BCI designs calculate *SCP* signals using low-pass filtering methods, and 64% of the *VEP*-based BCIs use *power-spectral* features at specific frequencies. 26% of the BCIs based on *P300* calculate the *PP*; 22% use *TFR*-based methods, 22% use no feature extraction method, and 15% use cross-correlation with a specific template. 41% of the BCI designs that use mental tasks to control a BCI use *power spectral* features and 37% use *parametric modelling* of the input signal. As most of the BCI designs that are based on neural

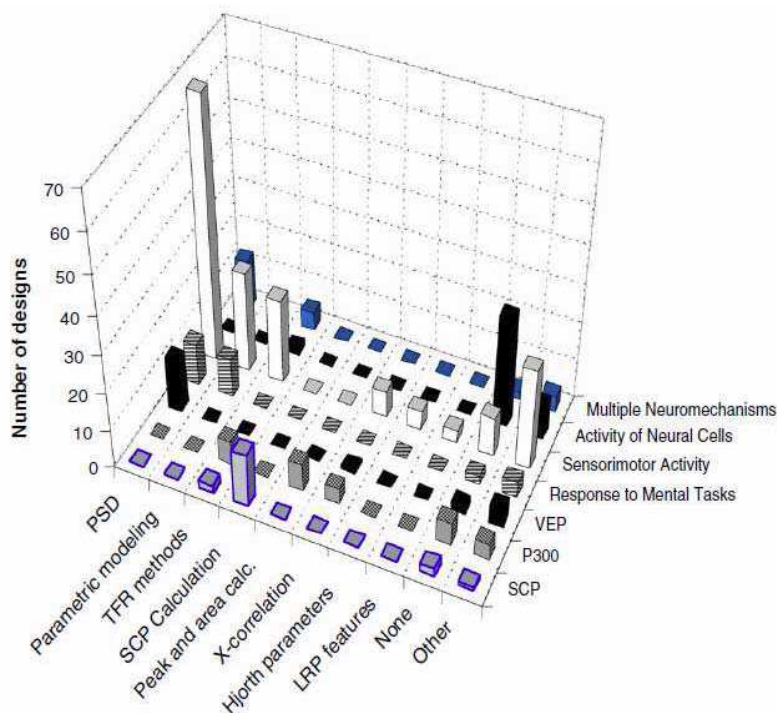


Fig. 6. Feature extraction methods in BCI designs based on sensorimotor activity, VEP, P300, SCP, response to mental tasks, activity of neural cells, and multiple neuromechanisms. Taken from (Bashashati et al., 2007).

cortical recordings mainly try to model the direct relationship between the neural cortical recordings and movements, they do not use a feature-extraction algorithm. 45% of the BCI designs that are based on multiple neuromechanisms rely on power-spectral features, 17% use parametric modelling, and 17% use TFR methods.

5. Feature selection and dimensionality reduction methods in BCI designs

In BCI applications, several features are generally extracted from several brain activity channels (several electrodes in the case of EEG measurements) and from several time segments (or sessions), before being concatenated into a single feature vector. Hence, the BCIs are often affected by a problem known as *curse of dimensionality* (Bellman, 1961). It was demonstrated that the amount of data needed to properly describe the different classes increases exponentially with the dimensionality of the feature vectors (Friedman, 1997; Jain, 2000).

(Flotzinger et al., 1994) and (Pfurtscheller & Guger, 1999b) have shown that when feature selection is used, the *classification* accuracy is better than when all the features are used. If the number of training data is small relatively to the number of features, the *classification* algorithm which will use these features and data will very likely give bad results. It is recommended to use at least 5 to 10 times more training data per class than the number of

features (Jain & Chandrasekaran, 1982; Raudys & Jain, 1991). Unfortunately this cannot be applied in all BCI systems as generally the dimensionality is high and the training set small. Among fourteen feature selection and dimensionality reduction methods in BCI designs given by (Bashashati et al., 2007), we give here briefly the definitions of only three methods which are the most applied in BCI designs:

- **Genetic algorithm (GA)**

A *Genetic algorithm* (GA) (Goldberg, 1989; Flotzinger et al., 1994) is a search technique used in computing to find exact or approximate solutions to optimization and search problems. Genetic algorithms are categorized as global search heuristics. They are a particular class of *evolutionary algorithms* (EA) that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover. These algorithms are based on a sequence of generations whereby the population in each generation produces the next while trying to optimize some *fitness criterion* (Brill, 1992), such as maximum ability to classify the training set in the *classification* stage. Each member of the current population is assigned a binary-valued chromosome of length n (for an n -dimensional *classification* problem), whereby the value of each bit within the chromosome defines whether this feature is to be used for *classification* or not. A chromosome 11 11 1 ...1, therefore, means that all parameters are to be used for evaluation of the member's fitness and a chromosome 10100 ... 0 means that only the first and third parameters are to be used. The accuracy which can be achieved using a specific chromosome is calculated using a *clustering algorithm* such as *k-nearest-neighbour classifier* (K-NN) or any other *clustering algorithm*. For *k-nearest-neighbour classification* the k nearest data vectors are found for a new input vector which is then classified to the label to which the majority of these k data vectors belong. For more details, refer to (Flotzinger et al., 1994; Brill, 1992).

- **Principal component analysis (PCA)**

Principal component analysis (PCA) may be used in *pre-processing* stage of BCI designs (see section 3.3). PCA may also be used as a dimensionality reduction technique in terms of capturing the variance of the data, and it accounts for correlation among variable. It gives lower-dimensional representations of the data which better generalize to data independent of the training set than using the entire dimensionality of the observation space (Scholkopf, 1999). The PCA transforms a set of m variables into another set of $k \leq m$ uncorrelated variables, maintaining as many of the variance of original data as possible (Moghaddam, 2002).

- **Distinctive sensitive learning vector quantization (DSLQV)**

The influence of distinctive features on the distance function in the standard *learning vector quantisation* (LVQ) (Kohonen, 1990) is equal. The *Distinction Sensitive* (DS) algorithm (DSLQV) (Flotzinger et al., 1994) employs an *adaptive weighted distance function* where the influence of features which frequently contribute to misclassifications is reduced while the influence of features which are shown to be very significant for proper *classification* is increased. For the *weighted distance function* of DSLQV a *global weights vector* w is used which stores the distinctiveness, i.e. the *relevance*, of every single feature. This weights vector is adapted interactively along with the codebook. The distance function used may be the *Euclidean distance*, or any other *weighted distance functions*. (Flotzinger et al., 1994) proposed the following *Euclidean distance* between two feature vectors x and y :

$$DSdist(x, y, w) = \sqrt{\sum_{i=1}^N (w_i(x_i - y_i))^2}. \quad (6)$$

The weights vector w can be seen as a scaling transformation from the original feature space into a DS -feature space. This transformation increases distances for very distinctive features and decreases distances for common features. Despite the usage of a *weighted distance function*, the codebook learning for $DSLQ$ is the same as for the $LVQ3$ algorithm. Additionally, the weights vector w must be updated with every learning iteration. Learning weights and codebook settings in parallel facilitate a quick approximation of these related parameters. For more details, refer to (Flotzinger et al., 1994).

Fig. 7 shows the statistical results of the study realised by (Bashashati et al., 2007) concerning *feature selection* and *dimensionality reduction* methods in *BCI* designs. (Bashashati et al., 2007) concluded that thirty-eight of the reported *BCI* designs employ feature selection and dimensionality reduction algorithms; 26% of these 38 designs use *genetic algorithms* (GA), 24% use *distinctive sensitive learning vector quantization* (DSLQ), and 13% use PCA.

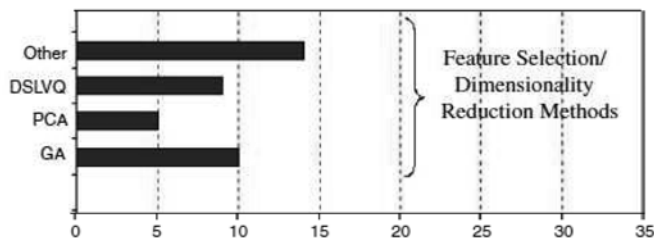


Fig. 7. Feature selection and dimensionality reduction methods in *BCI* designs. Figure modified from (Bashashati et al., 2007) with authorisation.

6. Classification in BCIs

Brain activity patterns are considered as dynamic stochastic processes due both to biological and to technical factors. Biologically, they change due to user fatigue and attention, due to disease progression, and with the process of training. Technically, they change due to amplifier noises, ambient noises, and the variation of electrode impedances (Wolpaw et al., 2002). Therefore, the time course of the generated time series signals (e.g. *EEG*) should be taken into account during feature extraction (Wolpaw et al., 2002). To use this temporal information, three main approaches have been proposed (Lotte et al., 2007):

- concatenation of features from different time segments: extracting features from several time segments and concatenating them into a single feature vector (Pfurtscheller et al., 1997; Haselsteiner & Pfurtscheller, 2000);
- combination of *classifications* at different time segments: it consists in performing the feature extraction and *classification* steps on several time segments and then combining the results of the different *classifiers* (Penny & Roberts, 1999; Lemm et al., 2004);
- dynamic *classification*: it consists in extracting features from several time segments to build a temporal sequence of feature vectors. This sequence can be classified using a dynamic *classifier* (Haselsteiner & Pfurtscheller, 2000; Obermeier et al., 2001).

The first approach is the most widely used despite that the obtained feature vectors are often of high dimensionality.

6.1 Classifier selection criteria

In order to choose the most appropriate *classifier* for a given set of features, the properties of the available *classifiers* must be chosen according to the following four *classifier* taxonomy and two main problems in BCI design as described by (Lotte et al., 2007):

6.1.1 Classifier taxonomy

- **Generative or Informative classifier - Discriminative classifier**
Generative classifiers, e.g., *Bayes quadratic*, learn the class models. To classify a feature vector, *generative classifiers* compute the likelihood of each class and choose the most likely. *Discriminative ones*, e.g., *support vector machines (SVM)*, only learn the way of discriminating the classes or the class membership in order to classify a feature vector directly (Ng & Jordan, 2002; Rubinstein & Hastie, 1997);
- **Static classifier - Dynamic classifier**
static classifiers, e.g., *multilayer perceptrons (MP)*, cannot take into account temporal information during *classification* as they classify a single feature vector. In contrast, *dynamic classifiers* such as *hidden Markov model (HMM)* (Rabiner, 1989), *FIR filters-multilayer perceptrons (FIR-MLP)* (Haselsteiner & Pfurtscheller, 2000) and *Tree-based neural network (TBNN)* (Ivanova et al., 1995), can classify a sequence of feature vectors and thus catch temporal dynamics.
- **Stable classifier - Unstable classifier**
Stable classifiers, e.g., *linear discriminant analysis (LDA)*, have a low complexity (or capacity (Vapnik, 1995; 1999)). They are said to be *stable* as small variations in the training set do not considerably affect their performance. In contrast, *unstable classifiers*, e.g., *multilayer perceptron*, have a high complexity. As for them, small variations of the training set may lead to important changes in performance (Breiman, 1998).
- **Regularized classifier**
Regularization consists in carefully controlling the complexity of a *classifier* in order to prevent *overtraining*. Regularization helps limit (a) the influence of *outliers* and strong *noise*, (b) the *complexity* of the *classifier* and (c) the *raggedness* of the decision surface (Müller et al., 2003). A *regularized classifier* has good *generalization* performances and is more robust with respect to *outliers* (Duda, 2001; Jain, 2000) (see section 6.1.2).

6.1.2 Main classification problems in BCI research

While performing a pattern recognition task, *classifiers* may be facing two main problems related to the feature properties. These problems are: the *curse-of-dimensionality* and the *bias-variance tradeoff*. The first problem is discussed in section 5. The second problem is a general one where three inherent errors may occur.

Independently of the applications of the *classification*, this stage consists in association of the true *class* label c^* corresponding to a feature vector x using a mapping (or a *model* f that may be learnt from a training set T . Let f^* the optimal unknown theoretical mapping that has generated the labels. If the *mean square error (MSE)* is considered to evaluate the inherent classification errors, then these errors can be decomposed into three terms (Breiman, 1998; Friedman, 1997):

$$\begin{aligned}
MSE &= E[(c^* - f(x))^2] \\
&= E[(c^* - f^*(x) + f^*(x) - E[f(x)] + E[f(x)] - f(x))^2] \\
&= E[(c^* - f^*(x))^2] + E[(f^*(x) - E[f(x)])^2] + E[(E[f(x)] - f(x))^2] \\
&= \text{noise}^2 + \text{bias}(f(x))^2 + \text{variance}(f(x)),
\end{aligned} \tag{7}$$

where the expectation is with respect to the distribution over c^* and the introduced term $E[f(x)]$ is the mean value of the estimated class label. These three terms describe three possible sources of inherent *classification* error:

- **noise**: represents the irreducible noise within the system;
- **bias**: represents a systematic error which is the divergence between the estimated mapping (i.e. the estimated class label) and the best mapping (i.e. the true class label). This term depends strongly on the classification method that has been chosen to obtain f (*linear, quadratic, ...*);
- **variance**: reflects the sensitivity to the used training set T .

The only terms that may be minimised are the *bias* and the *variance* terms. *Simple classifiers* (e.g. *linear classifiers*) have a high *bias* but low *variance* whereas more *complex classifiers* (e.g. *polynomial classifiers*) have a low *bias* but a high *variance*. To select the *optimal model complexity*, we must solve this *bias-variance dilemma* (Geman et al., 1992).

Actually, *stable classifiers* tend to have a high *bias* and a low *variance*, whereas the inverse is true for *unstable classifiers*. This can explain why *simple classifiers* sometimes outperform more complex ones. A low *variance* can be a solution to cope with the stochastic nature of brain signals. In the presence of strong noise and *outliers*, even linear systems can fail. One way of overcoming this problem is to use techniques, known as *stabilization techniques*. These techniques can be used to reduce the *variance*, e.g. *combination of classifiers* (Breiman, 1998) and *regularisation techniques* (Duda, 2001; Jain, 2000).

The *nonlinear classifiers* have more number of parameters than those in the *linear classifiers*. In addition, these parameters must be chosen appropriately. However, when there are large amounts and limited knowledge of data, *nonlinear classifiers* are better suited in finding the potentially more complex structure in the data. In particular, when the source of the data to be classified is not well known, using methods that are good at finding nonlinear transformation of the data is suggested. In these cases, *kernel-based* and *neural-networks* based methods can be used to determine the transformations. *Kernel-based classifiers* are classification methods that apply a linear classification in some appropriate (*kernel*) feature space. Thus, all the beneficial properties of linear classification are maintained, but at the same time, the overall classification is nonlinear. Examples of such *kernel-based* classification methods are *support vector machines (SVMs)* (Vapnik, 1999) and *kernel Fisher discriminant (KFD)* (Müller et al., 2003).

Better performances might also be achieved by using *group (committee)* of *classifiers* rather than using a single *classifier* but only a few BCI designs have employed such an approach in classifying features and achieved performance improvements (Millan et al., 2000; Peters et al., 2001; Varsta et al., 2000; Millan et al., 2002).

6.2 Classifiers used in BCI research

There are many categories of *classification* algorithms used to design BCIs. Among the more used *Classifiers* in the BCI community are: *linear discriminant classifiers (LDC)*, *neural networks (NN)*, *nonlinear Bayesian classifiers (NBC)*, *nearest neighbours classifiers (NNC)* and *combinations*

of classifiers. As a linear discriminant classifiers (LDC), we introduce in this section only the linear discriminant analysis (LDA) and support vector machines (SVMs), and as nonlinear Bayesian classifiers (NBC), we introduce in this section only the hidden Markov models (HMMs). For an extensive description and tutorials on these and other approaches, refer to (Lotte et al., 2007; Bashashati et al., 2007; Hung et al., 2005).

6.2.1 linear discriminant classifiers (LDC)

LDC are the most popular in BCI design. Two main kinds of LDC have been used for BCI design, namely, linear discriminant analysis (LDA) and support vector machine (SVM).

6.2.1.1 Linear discriminant analysis (LDA)

LDA or Fisher's LDA has been used with success in many of BCIs such as motor imagery based BCI (Pfurtscheller, 1999c), P300 applications (Bostanov, 2004; Hoffman et al., 2008), multiclass (Garrett et al., 2003) or asynchronous (Scherer et al., 2004) BCI. It has a very low computational requirement and it is simple to use and generally provides good results.

The idea of LDA (Fukunaga, 1990; Duda, 2001) is to find a weight vector \mathbf{w} so that two projected clusters c_1 and c_2 of N_1 and N_2 training feature vectors $\{x_1, x_2, \dots, x_{N_1}\}$ and $\{x_1, x_2, \dots, x_{N_2}\}$ on \mathbf{w} can be well separated from each other by hyperplanes while keeping small variance of each cluster. This can be done by maximizing the so-called Fisher's criterion. LDA assumes normal distribution of the data, with equal covariance matrix for both classes. The separating hyperplane is obtained by seeking the projection that maximizes the distance between the two classes' means and minimizes the between variance (Fukunaga, 1990).

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}}, \quad (8)$$

with respect to w , where S_b is the between-class variance matrix:

$$S_b = (m_{c_1} - m_{c_2})(m_{c_1} - m_{c_2})^T, \quad (9)$$

and S_w is the within-class variance matrix:

$$S_w = \sum_{x \in c_1} (x - m_{c_1})(x - m_{c_1})^T + \sum_{x \in c_2} (x - m_{c_2})(x - m_{c_2})^T, \quad (10)$$

where m_{c_1} and m_{c_2} are the cluster mean of classes c_1 and c_2 respectively.

The optimal weight vector \mathbf{w} is the eigenvector corresponding to the largest eigenvalue of $S_w^{-1} S_b$. Once this vector is obtained using the training data x_i is obtained by means of the training data, the classification then may be done by projecting the test feature vectors on it, and then the projected test feature vectors may be classified by employing a clustering rule such as k -nearest-neighbor decision rule.

In the case of multiclass separation problem ($c_i, i > 2$), several hyperplanes are used. The strategy generally used in this case is the one versus the rest (OVR) (Schlögl et al., 2005) which separate each class from all the others. This technique is suitable for the on-line BCIs because it has a very low computational requirement. It is simple to use and generally provides good results. The main drawback of LDA is its linearity that can provide poor results on complex nonlinear EEG data (Garcia et al., 2003). A regularized Fisher's LDA (RFLDA) has also been

used in the field of *BCI* (Blankertz et al., 2002; Müller et al., 2004) and give better results for *BCI* than the non-regularized version (Blankertz et al., 2002; Müller et al., 2004).

6.2.1.2 Support vector machine (SVM)

The generalization of the pattern recognition or regression results obtained from a limited sample constitutes the essential stake of the artificial training (machine learning). It is known that the only minimization of the empirical risk (the error of training) does not guarantee a weak error on a corpus of test. Thus the techniques of regularization, used since the years 1960, allow to carrying out a compromise between the capacity of the model to be learned (related to its complexity) and its aptitude to be generalized. From the point of view conceptual, the concept of structural risk introduced by Vladimir Vapnik into the years 1990 (Vapnik, 1995; Cortes & Vapnik, 1995; Vapnik, 1999) gives a bound of the error of test according to the error of training and of the complexity of the model. *Support vector machine* (SVM) have been applied in various domains for classification and regression.

In the following, we will outline briefly the idea of *SVM* for *binary classification* and then how this method can be applied to solve *multi-class classification* problems.

- **Binary classification SVM**

Support vector machines were originally designed for binary classification (Cortes & Vapnik, 1995). The main idea of this approach is to construct a *hyperplane* in order to separate two classes ($ci \in \{-1, 1\}$ or $ci \in \{0, 1\}$) so that the margin (Friedman, 1996) (the distance between the *hyperplane* and the nearest point(s) of each classes) is maximal and with minimal error. The theoretical principle of the *SVM* comprises two fundamental steps:

1. Make a nonlinear mapping Φ of the feature vectors from *input space* towards another space (*redescription space*) of higher dimension provided with a scalar product (*Hilbert space*), Fig. 8.

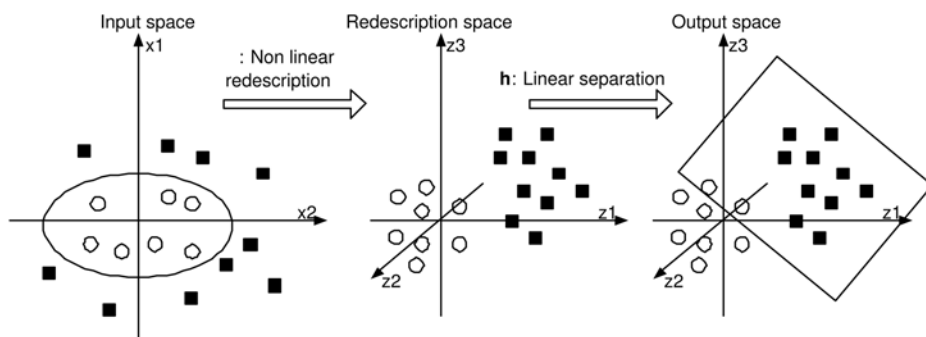


Fig. 8. The idea of the *SVM*: transform a nonlinear problem of separation in the *representation space* into a linear problem of separation in a *redescription space* of higher dimension.

2. Construct an optimal *hyperplane* allowing an optimal linear separation in this *redescription space*.

The interest of this mapping is that in the *redescription space* the pattern recognition can prove easy task: intuitively, more the dimension of the *redescription space* is high; more the probability of being able to find a *hyperplane* separating between the *feature vectors* is high, Fig. 9.

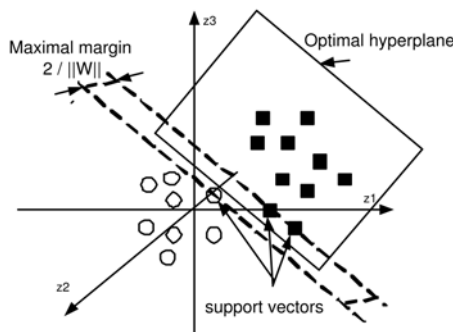


Fig. 9. Linear mapping: the optimal separating *hyperplane* is that which maximizes the margin in the *redescription space*.

From the mathematical point of view, the nonlinear mapping ($\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^n$) is realized via a *Kernel function* (*Hilbert-Schmidt Kernel*) easy to calculate. In practice, some families of Kernel functions are known and it is allocated to the user of SVM to carry out test to determine that which is appropriate for its application (it is a question of translating the maximum a priori knowledge of which one lays out on the studied problem and the data). Fig. 10 shows a non-linear separation example.

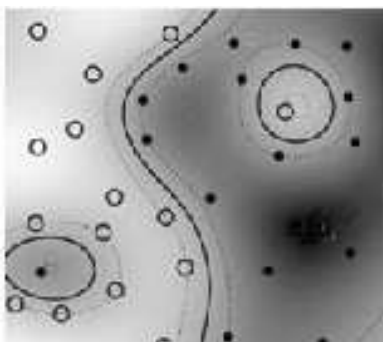


Fig. 10. *Non-linear mapping*: a complex shape. Figure from Schölkopf & Smola (2002).

Let \mathbf{x} denote a vector drawn from the *input space*, assumed to be of dimension N_0 and let $\{\Phi_j(\mathbf{x})\}, j = 1, 2, \dots, N_1$ denote a set of *nonlinear transformations* from the *input space* to the *redescription space*, where N_1 is the dimension of the *redescription space*. Given such a set of *nonlinear transformations*, we may define a *hyperplane* acting as the *decision surface* which is computed in the *redescription space* in terms of the *linear weights* of the machine as follows:

$$\sum_{j=0}^{N_1} \mathbf{w}_j \Phi_j(\mathbf{x}) = 0, \quad (11)$$

where $\mathbf{w} = \{\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{N_1}\}$ denotes a set of *linear weights* connecting the *redescription space* to the *output space*. And it is assumed that $\Phi_0(\mathbf{x}) = 1$ for all \mathbf{x} , so that \mathbf{w}_0 denotes the

bias. Define the vector $\Phi(\mathbf{x}) = [\Phi_0(\mathbf{x}), \Phi_1(\mathbf{x}), \dots, \Phi_{N_1}(\mathbf{x})]^T$, then we can rewrite the *decision surface* in the compact form:

$$\mathbf{w}^T \Phi(\mathbf{x}) = 0. \quad (12)$$

Given the *training feature vectors*, $\Phi(\mathbf{x}_i)$ corresponds to the *input pattern* \mathbf{x}_i , and the corresponding desired response d_i , $i = 1, 2, \dots, N_0$, which is either 1 or -1 (or, 0 or 1), it has been shown that (Haykin, 1994) the *optimal weight vector* \mathbf{w} can be expressed as

$$\mathbf{w} = \sum_{i=1}^{N_0} \alpha_i d_i \Phi(\mathbf{x}_i), \quad (13)$$

where $\{\alpha_i\}$, $i = 1, 2, \dots, N_0$ is the *optimal Lagrange multipliers* resulted from maximizing the *subject function*

$$Q(\alpha) = \sum_{i=1}^{N_0} \alpha_i - \frac{1}{2} \sum_{i=1}^{N_0} \sum_{j=1}^{N_0} \alpha_i \alpha_j d_i d_j \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}_j), \quad (14)$$

subject to the constraints

$$(1) \sum_{i=1}^{N_0} \alpha_i d_i = 0, \text{ and}$$

$$(2) 0 \leq \alpha_i \leq C,$$

where C is a user-specified constant which may be considered as *regulation parameter*. Substituting equation (13) into (12), we obtain the *optimal hyperplane*

$$\sum_{i=1}^{N_0} \alpha_i d_i \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}) = 0, \quad (15)$$

which will be used for linearly separating the *testing data*, i.e. for any testing sample \mathbf{x} , if

$$\sum_{i=1}^{N_0} \alpha_i d_i \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}) \geq 0, \quad (16)$$

then \mathbf{x} is classified into the subset having the training response $d_i = 1$, otherwise it is classified into the other subset with $d_i = -1$ (or 0). The user may choose the *radial basis function* (RBF) in defining the *inner product kernel* $K(\mathbf{x}, \mathbf{x}_i) = \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x})$.

According to equation (16), once the number of nonzero Lagrange multipliers, α_i , is determined, the number of *radial basis functions* and their centers are determined automatically. This differs from the design of the conventional neural network, such as the *back-propagation neural network* or *radial-basis function network* (Haykin, 1994), where the numbers of hidden layers or of hidden neuron are usually determined heuristically. The *kernel* generally used in BCI research is the *Gaussian* or *radial basis function* (RBF) kernel:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right). \quad (17)$$

In this case, the SVM is known as *Gaussian SVM* or *RBF SVM* (Burges, 1998; Bennett & Campbell, 2000; Garrett et al., 2003; Kaper et al., 2004).

This classifier has been applied, with success, to a relatively large number of BCI problems (Blankertz et al., 2002; Garrett et al., 2003; Rakotomamonjy et al., 2005; Helmy et al., 2008; Trad et al., 2009).

- Multiclass SVM (M-SVM)

To solve a given multiclass problem it is preferable to use a combination of several *binary SVM classifiers* and a decision strategy to decide the class of the input pattern. The most popular methods to achieve this goal are: *one-versus-all* method using *Winner-Takes-All* strategy (WTA-SVM); *One-Versus-One* method implemented by *Max-Wins Voting* (MWV-SVM) (Duan et al., 2007); *Decision Directed Acyclic Graph* (DAGSVM) (Platt et al., 2000); and *error correcting codes* (Dietterich & Bakiri, 1995).

6.2.2 Nonlinear Bayesian Classifiers (NBC)

This section introduces one Bayesian classifier used for BCI: *hidden Markov models* (HMMs). This Classifier produces *nonlinear decision boundaries*. Furthermore, they are *generative*, which enables them to perform more efficient rejection of uncertain samples than *discriminative classifiers*. As brain signals used to drive BCI have specific time courses, HMM have been applied to the classification of temporal sequences of BCI features or even raw EEG (Obermeier et al., 2000; 2001; Cincotti et al., 2003; Schlogl et al., 2005; Helmy et al., 2008; Trad et al., 2009).

HMMs are very efficient nonlinear techniques used for the classification of time series (Rabiner, 1989). HMMs have been widely applied in automatic speech recognition and now being applied to many fields, e.g. signal processing, pattern recognition, modelling and control. Their use necessitates two stages: a *training stage* where the stochastic process models are estimated through extensive observation corpus and decoding or *detection stage* where the model may be used off/on-line to obtain the likelihoods of the given testing sequence evaluated by each model.

In its conventional definition, a HMM is defined by the following compact notation to indicate the complete parameter set of the model $\lambda = (\Pi, \mathbf{A}, \mathbf{B})$, where Π , \mathbf{A} and \mathbf{B} are the initial state distribution vector, matrix of state transition probabilities and the set of the observation probability distribution in each state, respectively: $\Pi = [\pi_1, \pi_2, \dots, \pi_N]$, with $\pi_i = P(q_1 = s_i)$, $\mathbf{A} = [a_{ij}]$, with $a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$, $1 \leq i, j \leq N$, $s_i, s_j \in S$, $S = \{s_1, s_2, \dots, s_N\}$, $t \in \{1, 2, \dots, T\}$. The observation at time (or index) t , \mathbf{O}_t , is considered in this paper as continuous $\mathbf{O}_t \in \mathbb{R}^K$. For a continuous observation (CHMMs case), the state conditional probability of the observation $b_i(\mathbf{O}_t)$ may be defined by a finite mixture of any log-concave or *elliptically symmetric probability density function (pdf)*, e.g. *Gaussian pdf*, with *state conditional observation mean vector* $\bar{\mathbf{a}}_i$ and *state conditional observation covariance matrix* Σ_i . In this paper we consider only a single *Gaussian pdf*, so \mathbf{B} may be defined as $\mathbf{B} = \{\bar{\mathbf{a}}_i, \Sigma_i\}$, $i = 1, 2, \dots, N$. At each instant of time t , the model is in one of the states i , $1 \leq i \leq N$. It outputs \mathbf{O}_t according to a *density function* $b_i(\mathbf{O}_t)$ and then jumps to state j , $1 \leq j \leq N$ with probability a_{ij} . The *state transition matrix* defines the structure of the HMM (Rabiner, 1989). The model λ may be obtained off-line by a given training procedure. In practice, given the observation sequence $\mathbf{O}_l = \{\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_l\}$, $1 \leq l \leq L$ and a model λ , the HMMs need three fundamental problems to be solved

1. How to calculate the likelihood $P(O_i | \lambda)$? The solution to this problem provides a score of how O_i belongs to λ .
2. How to determine the most likely state sequence that corresponds to O_i ? The solution to this problem provides the sequence of the hidden states corresponding to the given observation sequence O_i .
3. How to adjust the model λ in order to maximize $P(O_i | \lambda)$? This is the problem of estimating the model parameters given a corpus of training observations sequences.

Problems 1 and 2 are solved in the decoding or detection stage using the forward or the *Viterbi algorithms* (Rabiner, 1989), while problem 3 is solved during the training phase using either a conventional algorithm such as the Baum-Welch algorithm (Rabiner, 1989) or another optimization based algorithm such as the simulated annealing based algorithm (Alani & Hamam, 2010).

7. Hidden Markov models and support vector machines approaches applied to P300-based Brain-Computer interface - a comparative study

This section reports on our contribution to investigate the use of *Hidden Markov Models* (HMMs) and *Support Vector Machines* (SVMs) approaches for tasks classification in P300-based Brain-Computer Interface (BCI) system. These approaches are applied to electroencephalogram (EEG) data sets of the École Polytechnique Fédérale de Laussane - BCI Groupe. These data are issued from five disabled subjects and four able-bodied subjects. For each subject, the HMMs and SVM are trained on different sets. Each set represents a specific P300 *potential evoked* task and contains multiple sequences (issued from different trials) of EEG records. For each task, the HMMs and the SVM that have been built take into account the variability of EEGs during different trials. Based on *Bayesian Inference Criterion* (BIC) (Schwarz, 1978), the proposed HMM training algorithm is able to select the optimal number of states corresponding to each task. This scheme makes the training procedure independent of the initialization problem and the a priori knowledge of the optimal number of HMM-related states needed in the training algorithm. The SVM is trained by introducing the different sets at the same time. We report training procedures and testing results for the HMMs and the SVM then we compare the performance of the two approaches on the same testing data sets. We also present promising preliminary comparative results based on the two classification approaches with correct classification rates for all subjects.

7.1 Problem definition

It has been shown by (Hoffman et al., 2008) that evoked potential (P300) in EEG data due to different images (tasks) appearing randomly on a computer screen in face of the subject may be efficiently used for exogenous mental task recognition. In Hoffman's work, extensions of LDA have been employed for pattern recognition. In our work we investigate the use of HMMs and SVM to classify the same P300 based tasks defined in (Hoffman et al., 2008) and using tasks data given in (Hoffman et al., 2009).

7.2 Data structure and feature extraction

For the clarity of the following sections, this section introduce briefly the experimental setup and the data given by Hoffman et al. (Hoffman et al., 2008), (Hoffman et al., 2009) that we adopted in our work.

7.3 Hoffman's experimental setup

In the experimental setup given by Hoffman et al., the users used a laptop screen displaying randomly six images (a television, a telephone, a lamp, a door, a window, and a radio), Fig. 11. Subjects were asked to count silently how often a prescribed image was flashed. The six images (Tasks) were displayed on the screen and a warning tone was issued.



Fig. 11. The display used for evoking the *P300*. One image was flashed at a time (Hoffman et al., 2008), (Hoffman et al., 2009).

These images were selected in order to control some electrical appliances based on the *BCI* system. Each image flashed during 100 ms followed by a period of 300 ms of non flashing. The sampling rate of *EEG* was 2048 Hz using 32 electrodes according to the 10-20 international system.

7.3.1 Subjects

The recorded data given by Hoffman et al. (Hoffman et al., 2009) and used in our work are issued from five disabled subjects (subjects 1 to 5) and four able-bodied subjects (subjects 6 to 9). The disabled subjects were all wheelchair-bound but had varying communication and limb muscle control abilities. Subjects 1 and 2 were able to perform simple, slow movements with their arms and hands but they were unable to control other extremities. Spoken communication with subjects 1 and 2 was possible, although both subjects suffered from mild dysarthria. Subject 3 was able to perform restricted movements with his left hand but

was unable to move his arms or other extremities. Spoken communication with subject 3 was impossible. However the patient was able to answer yes/no questions with eye blinks. Subject 4 had very little control over arm and hand movements. Spoken communication was possible with subject 4, although a mild dysarthria existed. Subject 5 was only able to perform extremely slow and relatively uncontrolled movements with hands and arms. Due to a severe hypophony and large fluctuations in the level of alertness, communication with subject 5 was very difficult, thus this subject was not taken into account by Hoffman et al. (Hoffman et al., 2008), (Hoffman et al., 2009). Subjects 6 to 9 were normal persons (all male, age 30 to 32). None of subjects 6 to 9 had known neurological deficits. In this study we use the data from all subjects.

7.3.2 Preprocessing and feature extraction

The data were preprocessed (referencing, bandpass-filtering, downsampling, and artefacts rejection). The *EEG* was downsampled from 2048 Hz to 32 Hz by selecting each 64th sample from the bandpass-filtered data. For each subject, the data were obtained during four sessions. One session comprised on average 810 trials, and the whole data for one subject consisted on average of 3240 trials. One trial duration takes 1000 ms. Single trials of duration 1000 ms were extracted from the data. Single trials started at stimulus onset, i.e. at the beginning of the intensification of an image, and ended 1000 ms after stimulus onset. Due to the Inter-stimulus interval of 400 ms, the last 600 ms of each trial were overlapped with the first 600 ms of the following trial.

7.3.3 Electrode selection

Hoffman et al. (Hoffman et al., 2008), (Hoffman et al., 2009) proved that the configuration with 8 electrodes gives the best result for the studying data Fig. 12.

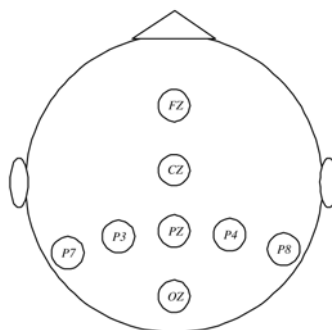


Fig. 12. The best configuration with 8 electrodes given by Hoffman et al. (Hoffman et al., 2008), (Hoffman et al., 2009).

7.3.4 Feature vectors construction

The samples from the selected electrodes were concatenated to obtain matrices of feature vectors. Each matrix is composed of N_e lines and N_t columns, where N_e denotes the number of electrodes and N_t denotes the number of temporal samples (feature vectors) in one trial. Due to the trial duration of 1000 ms and the downsampling to 32 Hz in (Hoffman et al., 2008) and (Hoffman et al., 2009), N_t is always taken equal to 32. In this work, we consider only

matrices of dimension 8×32 , i.e. each matrix O contains a sequence of 32 feature vectors $O = \{O_1, O_2, \dots, O_{32}\}$, $O_i \in \mathbb{R}^K$, $K = 8$. For the purpose of our *HMMs* or *SVM* training and testing schemes, we divided the set of all matrices corresponding to each subject into two parts: training set and testing set. Each set contains six subsets of multiple observation sequences (feature matrices $\{O_1, O_2, \dots, O_L\}$, $L \geq 1$), one subset for each image (Task) issued from different trials. Fig. 13 summarizes the training and data structures used in this work

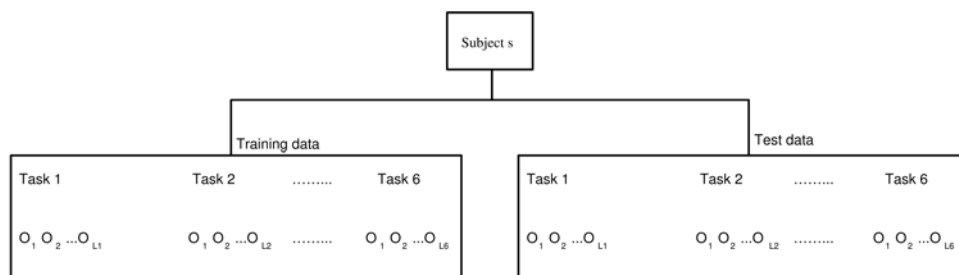


Fig. 13. Structures of the training and testing data sets for subject s , $1 \leq s \leq 8$. There are six subsets of multiple observation sequences (feature matrices $\{O_1, O_2, \dots, O_L\}$, $L \geq 1$), one subset for each image (Task). Each matrix is of dimension 8×32 .

7.4 Method

In this section, we give the *HMM* and *SVM* training schemes and their implementations in the context of our work.

7.5 HMM training algorithm

Given a set of L training data sequences $\{O_1, O_2, \dots, O_L\}$, $O_l = \{O_{l1}, O_{l2}, \dots, O_{lT_l}\}$, $1 \leq l \leq L$ issued from one of Z classes of stochastic processes p_z , $z \in 1, 2, 3, \dots, Z$. The problem of the training in our work is to build Z *HMMs* for each process p_z .

We consider that each *HMM* contains an optimal state number N_{opt} of internal states to be determined automatically between N_{min} and N_{max} . In order to determine N_{opt} , we developed an optimal state number selection algorithm based on the Bayesian Inference Criterion. This algorithm constructs a number of *HMMs* with different number of states between N_{min} and N_{max} given by the user. For every state number, each iteration is initialized by the most appropriate model using data clustering, and by the rejection of the least linked state or the rejection of the least probable state of the previous iteration. Consequently, every training iteration begins by a more precise model. After constructing these *HMMs*, the algorithm selects the optimal *HMM* with the number of states that maximize the *BIC*.

Given N_{min} and N_{max} , the algorithm may be summarized as follows:

1. Initialization (clustering): estimate the initial *HMM* λ_0 with N_{max} states using k-mean clustering algorithm.
2. While $N \geq N_{min}$, do:
 - a. run a training algorithm such a Baum-Welch algorithm (used in this work) until some convergence criterion is satisfied
 - b. calculate and save the selection criterion, $BIC(N)$, of the current model

$$BIC(N) = \log P(O|\lambda_N) - \frac{F_N}{2} \log(T_t).$$

F_N is the total number of free parameters of the current estimated *HMM* (λ_N) and T_t be the total number of observations in all training observation sequences, i.e. $T_t = \sum_{i=1}^L T_i$

- c. determine the number of zeros in each line of the state transition matrix, A .
- d. select the best initialization method for the next *HMM* according to the following criterion:
 - i. if all the lines of A contain zeros, then make a clustering of the observation data.
 - ii. if all the lines of A don't contain zeros but only some lines, then find and reject the state with minimal connections to other states.
 - iii. else, calculate the least probable state and reject it.
- e. obtain a reduced model $\bar{\lambda}$
- f. Fixe $N \leftarrow N - 1$ and $\lambda_N \leftarrow \bar{\lambda}$ and go to step 2

In this work we consider only a single *Gaussian pdf*, so \mathbf{B} may be defined as $\mathbf{B} = \{\bar{\sigma}_i, \Sigma_i\}$, $i = 1, 2, \dots, N$. In this case F_N is calculated by

$$F_N = N(N-1) + N[K + (K(K+1))/2] + (N-1),$$

where $\mathbf{O}_t \in \mathbb{R}^K$. In our case, we used $N_{\min} = 2$ and $N_{\max} = 4$. These values were chosen in order to not constraint the clustering algorithm because the number of feature vectors is always equal to 32 (see section 7.2). The training and testing data sets for each subject described in section 7.2 were used for the following training and testing schemes:

7.5.1 Training scheme

Using the above algorithm, the goal of this scheme is to construct, for each subject, six *HMMs*: *HMM1*, *HMM2*, ..., *HMM6* corresponding to six images (Tasks). Each *HMM* is constructed with either 2 states, 3 states or 4 states. The training data sequences for all subjects $s, 1 \leq s \leq 8$ and for the six images, all collected from different trials (see section 7.2), are given by the set $\{Tr_{s,l}\}$, $1 \leq s \leq 8$, $1 \leq l \leq 6$, where *Tr.* and *l.* denote training and image (or task) number respectively.

7.5.2 Off-line testing scheme

Our task is to classify six given testing feature sequences into class l , $1 \leq l \leq 6$. In order to make off-line (or on-line detection in future work) of these sequences, a stochastic dynamic programming based (*Viterbi decoding algorithm*) (Rabiner, 1989) has been used. This algorithm gives at the same time the solution to problems 1 and 2 of *HMMs* discussed in section 6.2.2. In this case, the likelihoods of the data sequence are compared with respect to all six built *HMMs*. The selected *HMM* (i.e. task) is the model that maximize the likelihood. The testing data sequences for all subjects $s, 1 \leq s \leq 8$ and for the six images, all collected from six trials (Hoffman et al., 2008), (Hoffman et al., 2009), are given by the set $\{Te_{s,l}\}$, $1 \leq s \leq 8$, $1 \leq l \leq 6$, where *Te* and *l* denote testing and image (or task) number respectively. Fig 14 show the training and testing schemes for one subject and for six images.

7.5.3 Multiclass SVM (M-SVM)

In our work, we used a modified version of the *multiclass SVM* (M-SVM). The M-SVM method adopted in this work is the *one-versus-all*. This method is done by winner takes-all

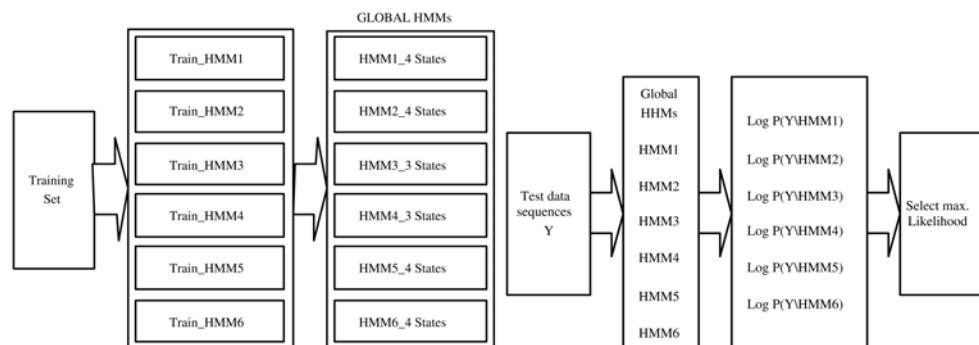


Fig. 14. Left: Training scheme for one subject and for six images. Right: Testing scheme for one subject given one or more task-related sequences.

combination strategy (Friedman, 1996), the outputs of the *binary one-vs-all classifiers* are compared and the most probable output among the other ones is chosen. The i^{th} SVM is trained with all of the examples in the i^{th} class with positive labels, and the rest with negative labels. An optimal hyperplane is constructed to separate N/M positive examples from $N(M - 1)/M$ negative examples where M is the number of SVMs and N is the number of the training examples used to construct an SVM for a given class. The same training and testing data structures, described in sections 7.2, 7.5.1 and 7.5.2, were used for SVM training and testing. To train or to test the SVM for a one subject, a number of $6 \times L$ training or testing feature matrices for all 6 tasks corresponding to this subject were concatenated and fed together to the M -SVM algorithm.

7.6 Results

The performance of BCI depends on different evaluation parameters. Among these evaluation parameters of interest in this work is the classification accuracy. This section reports experimental evaluations of the classification accuracy based on the testing scheme introduced in section 7.5.2. Using 60 testing feature matrices for all tasks corresponding to each subject (10 feature matrices for each task), the obtained classification results give the highest percentages of classification rates for the two applied approaches SVM and HMMs. The highest classification percent rates for all subjects are shown in the diagonals of the confusion matrices given in Table 1, see (Appendix). To conclude on the obtained results, we consider the case of comparison between the one of the more severely disabled subject, i.e. subject 3 (late-stage *amyotrophic lateral sclerosis*) and one of the able-bodied subjects, i.e. subject 6 (subject had no known neurological deficits). Fig. 15 show the HMMs-SVM based comparative result for subject 3 and for subject 6. Fig. 15 and Table 1 show that the overall classification percent rates give a satisfactory and promising results for all the disabled subjects compared to the able-bodied ones. Fig. 16 show the HMMs-SVM based comparative results of the means and standard deviations of all the tasks generated by all subjects.

Fig. 16 show that

- In both the HMMs and the SVM cases, the means and the standard deviations of the classification percent rates in the disabled subjects are approximately the same or slightly higher than those in the able-bodied.

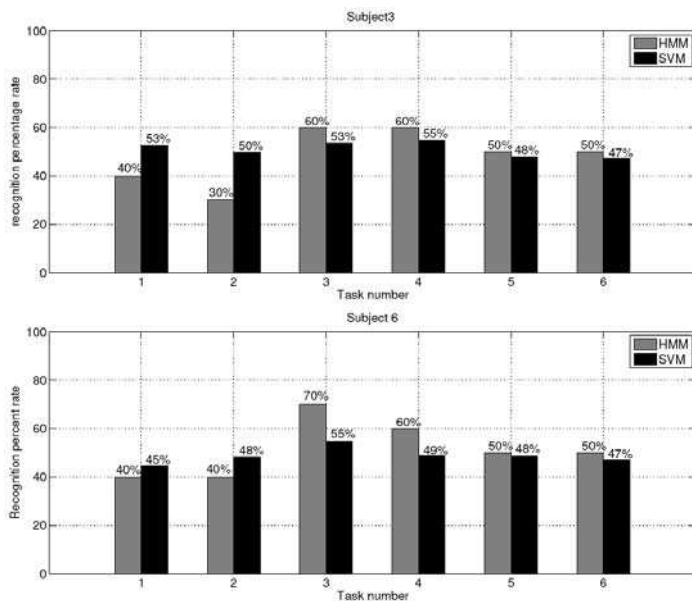


Fig. 15. Classification percent rates versus task number for subject 3 (top): late-stage amyotrophic lateral sclerosis and for subject 6 (down): subject had no known neurological deficits.

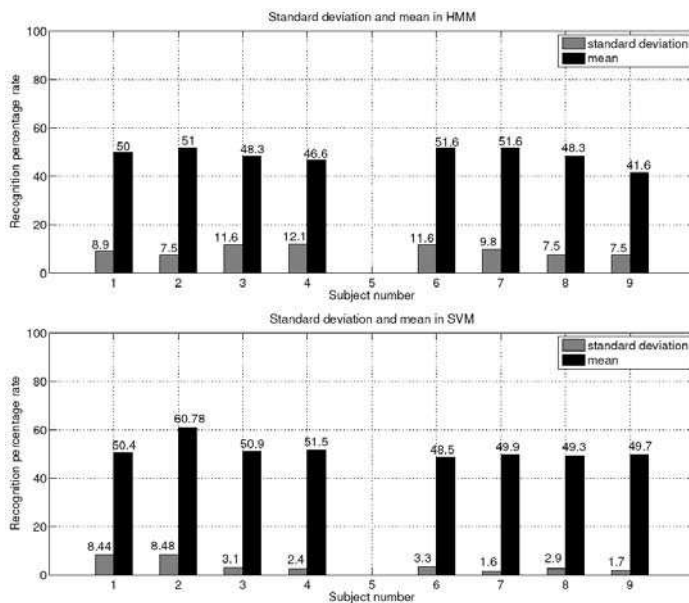


Fig. 16. Means and standard deviations of the classification percent rates in all tasks versus subject number (top: HMMs case, down: SVM case).

- In both the *HMMs* and the *SVM* cases, the means of the classification percent rates for the disabled subjects (except subject 2) as well as for the able-bodied subjects (except subject 9) are approximately uniformly distributed.
- The standard deviations of the classification percent rates for all subjects in the *SVM* case are less than or equal to those in the *HMMs* case.

These results give a way to study the ability of the different persons to learn the different *P300*-based *BCI* tasks.

8. Conclusions

This chapter have introduced a comprehensive survey of *signal processing*, *feature extraction/selection* and *classification* methods used to provide the readers with guidelines on design *brain-computer interfaces (BCIs)*. Based on the results given by (Bashashati et al., 2007) and (Lotte et al., 2007), this chapter shows which method have received more attention by the *BCI* community prior to 2006. However, other existing methods, not currently applied in *BCI* design, could be explored. The exploration of new methods would be strongly driven by the new properties that will have to be taken into consideration in the real future applications of *BCIs*.

As a contribution to the *classification* methods, a *Hidden Markov Models (HMMs)* and *support vector machines (SVM)* approaches for task *classification* in *P300*-based *BCI* system is presented. In the *HMMs* case, we proposed a training algorithm which is able to select automatically the optimal number of *HMM*-related states corresponding to each set of *EEG* training records. In the *SVM* case, we applied a *multiclass SVM* based on the *one-versus-all* method. The *confusion matrices* give a correct classification rates for the able-bodied subjects as well as for the disabled subjects. The comparative results demonstrate that the two approaches are promising. If the *HMMs* are not constructed correctly or the training data set is not sufficient, the *HMM* approach gives in this case a low classification rates. In such cases, a hybrid *HMM-SVM* may be employed, where the *HMM* is used as a *dynamic classifier* and the *SVM* is used as a good *discriminative classifier* by considering all the training examples in the training data set and train all the task-related models simultaneously. The authors are currently working on extending this work to include off-line and on-line training and classification schemes by using this strategy. The information transfer rate (bit rate) will be taken into account to include both accuracy and speed of a *BCI* system.

9. Acknowledgments

The authors would like to thank Dr. U. Hoffman and the EPFL-Brain-Computer team for the data and the software given in (Hoffman et al., 2008) that they were used in this work. The authors would like also to thank Dr. A. Bashashati for his authorization to use or modify some figures given in the paper (Bashashati et al., 2007) to illustrate some sections given in this chapter.

10. References

Anderson, C.W. & Sijercic, Z. (1996). Classification of EEG signals from four subjects during five mental tasks. In *Solving Engineering Problems with Neural Networks*.

- Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN.96).*
- Anderson, C.W.; Stolz, E. A. & Shamsunder, S. (1998). Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Trans Biomed Eng*, 45, pp. 277-286.
- Altenmüller, E. O.; Gerloff, C. (1999). Psychophysiology and the EEG. In: *Niedermeyer E, Lopes da Silva FH*, editors. *Electroencephalography: basic principles, clinical applications and related fields*, 4th ed. Baltimore, MD: Williams and Wilkins, pp. 637-655.
- Al-ani, T. & Hamam, Y. (2010). A low Complexity Simulated Annealing Approach For Training Hidden Markov Models. *Int. J. of Operational Research (IJOR)*, (Accepted for publication), Vol. 8, No. 3.
- Bayliss, J. D. & Ballard, D. H. (1999). Single trial P300 recognition in a virtual environment. *Proc. Int. ICSC Symp. on Soft Computing in Biomedicine*, (Genova, Italy).
- Bayliss, J. D. & Ballard, D. H. (2000a). Recognizing evoked potentials in a virtual environment. *Advances in Neural Information Processing Systems*, Vol 12, ed S A Solla, T K Leen and K R Müller (Cambridge, MA: MIT Press), pp. 3-9.
- Bayliss, J. D. & Ballard, D. H. (2000b). A virtual reality testbed for brain-computer interface research *IEEE Trans. on Rehabilitation Engineering*, Vol. 8, pp. 188-190.
- Bayliss, J. D.; Inverso, S. A. & Tentler, A. (2004). Changing the P300 Brain-Computer Interface. *Journal of CyberPsychology & Behavior*, Vol. 7, No.6, pp.694-704.
- Bayliss, J. D. & Inverso, S. A. (2005). Automatic Error Correction Using P3 Response Verification for a Brain-Computer Interface. *Proc. 11th International Conference on Human- Computer Interaction*, July 22-27, Las Vegas, NV. Mahwah: Lawrence Erlbaum Associates.
- Bashashati, A.; Ward R. K. & Birch, G. E. (2005). A new design of the asynchronous brain-computer interface using the knowledge of the path of features. *Proc. 2nd IEEE-EMBS Conf. on Neural Engineering*, (Arlington, VA), pp. 101-104.
- Bashashati, A.; Fatourehchi, M.; Ward R. K. & Birch, G. E. (2007). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals, *J. Neural Eng.*, June, Vol. 4, No. 2, pp. R32-R57.
- Bellman, R. (1961). *Adaptive Control Processes: A Guided Tour*. Princeton University Press, New Jersey.
- Bell, C. J.; Shenoy, P.; Chalodhorn, R. & Rao, R. P. N. (2008). Control of a humanoid robot by a noninvasive brain-computer interface in humans, *J. Neural Eng.* 5, pp. 214-220.
- Bennett, K. P. & Campbell, C. (2000). Support vector machines: hype or hallelujah?, *ACM SIGKDD Explor. Newslett.*, pp. 21-13.
- Birbaumer, N.; Flor, H.; Ghanayim, N.; Hinterberger, T.; Iversen, I.; Taub, E.; Kotchoubey, B.; Kubler, A. & Perelmouter, J. (1999). A spelling device for the paralyzed, *Nature*, Vol.398, No. 6725, pp. 297-298.
- Birbaumer, N.; Kübler, A.; Ghanayim, N.; Hinterberger, T.; Perelmouter, J.; Kaiser, J.; Iversen, I.; Kotchoubey, B.; Neumann, N. & Flor, H. (2000). The thought translation device (TTD) for completely paralyzed patients. *IEEE Trans. Rehab. Eng.*, Vol. 8, No. 2, pp. 190-192.

- Birbaumer, N. & Cohen, L. G. (2007). Brain-computer interfaces: communication and restoration of movement in paralysis. *The Journal of Physiology*, Vol. 579, No. 3, pp. 637-642.
- Blankertz, B.; Curio, G. & Müller, K. R. (2002). Classifying single trial EEG: towards brain computer interfacing. *Adv. Neural Inf. Process. Syst. (NIPS 01)*, Vol. 14, 157-164.
- Blankertz, B.; Dornhege, G.; Krauledat, M.; Müller, K. R.; Kunzmann, V.; Losch, F. & Curio, G. (2006). The Berlin brain-computer interface: EEG-based communication without subject training. *IEEE Trans. on Rehabilitation Engineering*, Vol. 14, No. 2., pp. 147-152.
- Boostani, R. & Moradi, M. H. (2004). A new approach in the BCI research based on fractal dimension as feature and adaboost as classifier. *Journal of Neural Engineering*, Vol. 1, No. 4, pp. 212-217.
- Bostanov, V. (2004). BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Trans. Biomed. Eng.*, Vol. 51, pp. 1057-1061.
- Breiman, L. (1998). Arcing classifiers. *Ann. Stat.*, Vol. 26, pp. 801-849.
- Brill, F.Z.; Brown, D.E. & Martin, W.N. (1992). Fast Genetic Selection of Features for Neural Network Classifiers. *IEEE Transactions on Neural Networks*, Vol. 3, pp. 324-328.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Knowl. Discov. Data Min.*, Vol. 2, pp. 121-167.
- Burke, D. P.; Kelly, S. P.; deChazal, P.; Reilly, R. B. & Finucane, C. (2005). A parametric feature extraction and classification strategy for brain-computer interfacing. *IEEE Trans. On Rehabilitation Engineering*, Vol. 13, No. 1.
- Cincotti, F.; Scipione, A.; Tiniperi, A.; Mattia, D.; Marciani, M. G.; Millán J. del R; Salinari, S.; Bianchi, L. & Babiloni, F. (2003). Comparison of different feature classifiers for brain computer interfaces. *Proc. 1st Int. IEEE EMBS Conf. on Neural Engineering*.
- Comon, P. (1994). Independent Component Analysis: a new concept?. *Signal Processing, Elsevier*, Vol. 36, No. 3, pp. 287-314.
- Congedo, M.; Lotte, F. & Lécuyer, A. (2006). Classification of movement intention by spatially filtered electromagnetic inverse solutions. *Physics in Medicine and Biology*, Vol. 51, No. 8, pp. 1971-1989.
- Cortes, C. & Vapnik, V. (1995). Support vector networks. *Machine Learning. Machine Learning*, Vol. 20, pp. 273-297.
- Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, Vol. 30, No. 1, pp. 34-43.
- Dietterich, T. & Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, Vol. 2, pp. 263-286.
- Dobkin, B. H. (2007). Brain-computer interface technology as a tool to augment plasticity and outcomes for neurological rehabilitation, *Journal of physiology*, vol. 579, No. 3, pp. 637-642.
- Dornhege, G.; Millán, J. del R; Hinterberger, T.; McFarland, D. J.; Müller, K-R; Sejnowski, T. J. (2007). An introduction to brain computer interfacing. In: *Towards. Brain-Computer Interfacing*, MIT Press, Cambridge.

- Duda, R. O.; Hart, P. E. & Stork, D. G. (2001). *Pattern Recognition. 2nd. ed. (New York: Wiley-Interscience).*
- Duan, K-B.; Rajapakse, J.C. & Nguyen, M.F. (2007). One-versus-one and one-versus-all multiclass SVM-RFE for gene selection in cancer classification. In: *EvoBIO*, pp. 228-234.
- Farwell, L. A. & Donchin, E. (1988). Talking off the top of your head: Towards a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.*, Vol. 80 510-523.
- Flotzinger, D.; Pregenzer, M. & Pfurtscheller, G. (1994). Feature selection with distinction sensitive learning vector quantisation and genetic algorithms. *Proc. IEEE Int. Conf. on Neural Networks*, (Orlando, FL), pp. 3448-3451
- Friedman, J. H. (1996). Another approach to polychotomous classification. *Technical report*, Department of Statistics, Stanford University.
- Friedman, J. H. K. (1997). On bias, variance, 0/1-loss, and the curse-of-dimensionality. *Data Min. Knowl. Discov.*, Vol. 1, pp. 55-77.
- Friedman, D.; Slater, M.; Steed, A.; Leeb, R.; Pfurtscheller, G. & Guger, C. (2004). Using a brain-computer-interface in a highly immersive virtual reality. In *IEEE VR Workshop*.
- Fukunaga, K. (1972). In: *Introduction to Statistical Pattern Recognition*, Oxford, UK., Clarendon.
- Fukunaga, K. (1990). *Statistical Pattern Recognition. 2nd edn (New York: Academic).*
- Garcia, G.; Ebrahimi, T. & Vesin, J. M. (2003). Support vector EEG classification in the Fourier and time-frequency correlation domains. *Proc. 1st IEEE-EMBS Conf. on Neural Engineering*, Capri Island, Italy, pp. 591-594.
- Garrett, D.; Peterson, D. A.; Anderson, C. W. & Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans. On Rehabilitation Engineering*, Vol. 11, pp. 141-144.
- Geman, S.; Bienenstock, E. & Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, Vol. 4, pp. 1-58.
- Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization. and Machine Learning. Addison-Wesley.*
- Grave de Peralta Menendez, Gonzalez Andino, R.; S.; Perez L.; Ferrez, P. & Millán, J. (2005). Non-invasive estimation of local field potentials for neuroprosthesis control. *Cognitive Processing*, Special Issue on Motor Planning in Humans and Neuroprosthesis Control.
- Gupta, S. & Singh, H. (1996) Preprocessing EEG signals for direct human-system interface. *Intelligence and Systems. IEEE International Joint Symposia on*, pp. 32-37.
- Guger, C.; Ramoser, H. & Pfurtscheller G. (2000). Real-Time EEG Analysis with Subject-Specific Spatial Patterns for a Brain-Computer Interface (BCI). *IEEE Trans. on rehabilitation engineering*, Vol. 8, No. 4, pp. 447-456.
- Guger, C.; Schlögl, A.; Neuper, C.; Walterspacher, D.; Strein, T. & Pfurtscheller, G. (2001). Rapid prototyping of an EEG-based brain-computer interface (BCI). *IEEE Trans. Rehab. Eng.*, Vol. 9, No. 1, pp. 49-58.

- Hagemann, D.; Naumann, E. & Thayer, J.F. (2001). The quest for the EEG reference revisited: A glance from brain asymmetry research. *Psychophysiology*, Vol. 38, 847-857.
- Haibin, Z.; Xu, W. & Hong, W. (2008). Brain-computer interface design using relative wavelet energy. *The 2nd International Conference on Control and Decision*, pp. 3558-3561, 2-4 July.
- Haselsteiner, E. & Pfurtscheller, G. (2000). Using time-dependant neural networks for EEG classification. *IEEE Trans. on Rehabilitation Engineering*, Vol. 8, pp. 457-463.
- Haykin, S. (1994). Neural Network: A Comprehensive Foundation. *New York: Macmillan College Publishing Company*.
- Helmy, S.; Al-ani, T.; Hamam, Y. & El-madbouly, E. (2008). P300 Based Brain-Computer Interface Using Hidden Markov Models. *ISSNIP 2008*, Sydney, 15-18 Dec.
- Hinterberger, T.; Kubler, A.; Kaiser, J.; Neumann, N. & Birbaumer, N. (2003). A brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device. *Electroencephalogr. Clin. Neurophysiol.* Vol. 114, pp. 416-25.
- Hinterberger, T.; Weiskopf, N.; Veit, R.; Wilhelm, B.; Betta, E. & Birbaumer, N. (2004). An EEGDriven Brain-Computer Interface Combined With Functional Magnetic Resonance Imaging (fMRI). *IEEE Transactions On Biomedical Engineering*, June, Vol. 51, No. 6.
- Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, Vol. 29, No. 3, pp. 306-310.
- Hoffmann, U.; Vesin, J.-M.; Ebrahimi, T. & Diserens K. (2008). An efficient P300-based braincomputer interface for disabled subjects, *Journal of Neuroscience Methods*, Vol. 167, No. 1, pp. 115-125.
- Hoffmann, U.; Vesin, J.-M.; Ebrahimi, T. & Diserens K. (2009). An efficient P300- based brain-computer interface for disabled subjects. Available from URL: <http://mmspl.epfl.ch/page33712.html>. Visited on September, 10, 2009.
- Huan, N.-J. & Palaniappan, R. (2004a). Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals. In: *IEMBS .04. 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1-5.
- Huang, N. E.; Shen, Z.; Long, S. R.; Wu, M. C.; Shih, H. H.; Zheng, Q.; Yen, N. C.; Tung, C. C. & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society*, Vol. 454, pp. 903-995.
- Huan, N.-J. & Palaniappan, R. (2004b). Neural network classification of autoregressive features from electroencephalogram signals for brain-computer interface design. *J. Neural Eng.*, Vol. 1, pp. 142-150.
- Hung, C.-I.; Lee, P.-L.; Wu, Y.-T.; Yeh, T.-C. & Hsieh, J.-C. (2005). Recognition of Motor Imagery Electroencephalography Using Independent Component Analysis and Machine Classifiers. *Annals of Biomedical Engineering*, Vol. 33, No. 8, pp. 1053-1070.
- Hyvärinen, A. & Oja, E. (2000). Independent Component Analysis: A Tutorial. *Neural Networks*, Vol. 13, No. 4-5, pp. 411-430.

- Ivanova, I.; Pfurtscheller, G. & Andrew, C. (1995). AI-based classification of single-trial EEG data. *Proc. 17th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, (Montreal, Canada), pp. 703704.
- Jain, A. K. & Chandrasekaran, B. (1982). Dimensionality and sample size considerations in pattern recognition practice. In: *Handbook of Statistics*, Vol. 2, pp. 835-855.
- Jain, A. K.; Duin, R. P.W. & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, pp. 4-37.
- Jung, T-P.; Humphries, C., Lee, T-W., Makeig, S.; McKeown, M. J.; Iragui, V. & Sejnowski T. J. (1998). Removing electroencephalographic artefacts: comparison between ICA and PCA. *Neural Networks Signal Processing*, Vol. III, pp. 63-72.
- Jung, T-P.; Makeig, S.; Westerfield, M.; Townsend, J.; Courchesne, E. & Sejnowski T. J. (2000). Removal of eye activity artefacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, Vol. 111, pp. 1745-1758.
- Kaper, M.; Meinicke, P.; Grosse-Kathoefer, U.; Lingner, T. & Ritter, H. (2004). BCI competition 2003-data set iib: support vector machines for the p300 speller paradigm. *IEEE Trans. Biomed. Eng.*, Vol. 51, pp. 1073-1076.
- Keirn, Z.A. & Aunon, J.I. (1990). A new mode of communication between man and his surroundings. *IEEE Trans Biomed Eng.*, 37, pp. 1209-1214.
- Kohonen, T. (1990). The Self-Organizing Map. *Proc. IEEE*. Vol. 78, No. 9, pp. 1161-1180.
- Koles, Z. J. (1991). The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG, *Electroenceph. Clin. Neurophysiol.*, Vol. 79, pp. 440-447.
- Lang, W.; Cheyne, D.; Hollinger, P.; Gerschlag, W. & Lindinger, G. (1996). Electric and magnetic fields of the brain accompanying internal simulation of movement, *Cogn Brain Res*, 3, pp. 125-129.
- Lehmann, E. & Skrandies, W. (1984). Spatial Analysis of Evoked Potentials in Man - A Review. *Progress in Neurobiology*, Vol. 23, pp. 227-250.
- Lee, H. & Choi, S. (2003). PCA+HMM+SVM for EEG pattern classification. In: *Proceedings of the Seventh International Symposium on Signal Processing and Its Applications*.
- Lemm, S.; Schafer, C. & Curio, G. (2004). BCI competition 2003-data set iii: probabilistic modelling of sensorimotor mu rhythms for classification of imaginary hand movements. *IEEE Trans. Biomed. Eng.*, Vol. 51, pp. 1077-1080.
- Lécuyer A.; Lotte, F.; Reilly, R. B.; Leeb, R.; Hirose, M.; Slater, M. (2008). Brain-Computer Interfaces, Virtual Reality, and Videogames. *IEEE Computer*, Vol. 41, No. 10, pp.66-72.
- Lotte, F.; Congedo, M.; Lécuyer, A.; Lamarche, F. & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.*, 4, pp. R1-R13.
- Mason, S.; Bashashati, A.; Fatourechi, M.; Navarro, K. & Birch, G. (2007). A comprehensive survey of brain interface technology designs, *Annals of Biomedical Engineering*, Vol. 35, No. 2, pp. 137-169.
- Meinicke, P.; Kaper, M.; Hoppe, F.; Heumann, M., & Ritter, H. (2003). Improving Transfer Rates in Brain Computer Interfacing: a Case Study. In: *Advances in Neural*

- Information Processing Systems*, Cambridge, MA, 15, Eds.: Suzanna Becker and Sebastian Thrun and Klaus Obermayer, MIT Press.
- McFarland, D. J.; McCane, L. M.; David, S. V. & Wolpaw, J. R. (1997a). Spatial filter selection for EEG-based communication Electroenceph. *Clin. Neurophysiol.*, Vol. 103, pp. 386-394.
- McFarland, D. J.; Lefkowitz, A. T. & Wolpaw, J. R. (1997b). Design and operation of an EEGbased brain-computer interface (BCI) with digital signal processing technology. *Behav. Res. Meth. Instrum. Comput.*, Vol. 29, pp. 337-345.
- McFarland D. J.; Miner L. A.; Vaughan T. M. & Wolpaw J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movement. *Brain Topogr*, 3, pp. 177- 186.
- McFarland, D. J. & Wolpaw, J. R. (2005). Sensorimotor rhythm-based brain-computer interface (BCI): feature selection by regression improves performance. *IEEE Trans. on Rehabilitation Engineering*, Vol. 13, No. 3, pp. 372-379.
- Millan, J.; Mourino, J.; Cincotti, F.; Varsta, M.; Heikkonen, J.; Topani, F.; Marciani, M. G.; Kaski, K. & Babiloni, F. (2000). Neural networks for robust classification of mental tasks. *Proc. 22nd Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, (Chicago, IL), pp. 1380-1382.
- Millan, J.; Mourino, J.; Franze, M.; Cincotti, F.; Varsta, M.; Heikkonen, J. & Babiloni, F. (2002). A local neural classifier for the recognition of EEG patterns associated to mental tasks. *IEEE Trans. Neural Netw.*, Vol. 13 pp. 678-686.
- Millán J. R. & Mouriño, J. (2003). Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project. *IEEE Trans. on Rehabilitation Engineering*, Vol. 11 159-61.
- Moghaddam, B. (2002). Principal Manifolds and Probabilistic Subspaces for Visual Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6, pp. 780-788.
- Müller-Gerking, J.; Pfurtscheller, G. & Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology*, Vol. 110, No. 5, pp. 787-798.
- Müller, K. R.; Anderson, C.W. & Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *IEEE Trans. on Rehabilitation Engineering*, Vol. 11, pp. 165-169.
- Musha, T.; Terasaki, Y.; Haque, H.A. & Ivanitsky, G.A. (1997). Feature extraction from EEGs associated with emotions. *Artificial Life and Robotics*, Vol. 1, No. 1, pp. 15-19.
- Müller, K. R.; Krauledat, M.; Dornhege, G.; Curio, G. & Blankertz, B. (2004). Machine learning techniques for brain-computer interfaces. *Biomed. Technol.*, Vol. 49, pp. 11-22.
- Nakamura, A.; Sugi T.; Ikeda, A.; Kakigi, R. & Shibasaki, H. (1996). Clinical application of automatic integrative interpretation of awake background EEG: quantitative interpretation, report making, and detection of artefacts and reduced vigilance level. *Electroenceph. Clin. Neurophysiol.*, Vol. 98, pp. 103-112.
- Neuper, C. & Pfurtscheller, G. (1999a). Motor imagery and ERD Related Desynchronization, *Handbook of Electroencephalography and Clinical Neurophysiology*, (Revised Edition), pp. 303-525, Vol. 6, Elsevier, Amsterdam.

- Neuper, C.; Schlögl, A. & Pfurtscheller, G. (1999b). Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery. *J. Clin. Neurophysiol.*, Vol. 16, pp. 373-382.
- Ng, A. Y. & Jordan, M. I. (2002). On generative versus discriminative classifiers: a comparison of logistic regression and naive Bayes. *Neural Information Processing Systems*, Ng, A.Y., and Jordan, M.
- Nicolelis, M.A.L. (2001). Actions from Thoughts. In: *Nature*, Vol. 409, January 18, pp. 403-407.
- Obermaier, B.; Neuper, C., Guger, C. & Pfurtscheller, G. (2000). Information transfer rate in a five-classes brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 9, pp. 283-288.
- Obermaier, B.; Guger, C.; Neuper, C. & Pfurtscheller, G. (2001). Hidden markov models for online classification of single trial EEG. *Pattern recognition letters*, pp. 1299-1309.
- Offner, F.F. (1950). The EEG as Potential Mapping: The Value of the Average Monopolar Reference. *Electroencephalography and Clinical Neurophysiology*, Vol. 2, pp. 215-216.
- Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* Vol. 2, No. 6, pp. 559- 572.
- Penny, W. D. & Roberts, S. J. (1999). EEG-based communication via dynamic neural network models. *Proc. Int. Joint Conf. on Neural Networks*.
- Peters, B. O.; Pfurtscheller, G. & Flyvbjerg, H. (2001). Automatic differentiation of multichannel EEG signals. *IEEE Trans. Biomed. Eng.*, Vol. 48, pp. 111-116.
- Pfurtscheller, G.; Kalcher J.; Neuper, C.; Flotzinger D & Pregenzer, M. (1996). On-line EEG classification during externally-paced hand movements using a neural network-based classifier. *Electroenceph. Clin. Neurophysiol.*, Vol. 99, pp. 416-425.
- Pfurtscheller, G.; Neuper, C.; Flotzinger D & Pregenzer, M. (1997). EEG-based discrimination between imagination of right and left hand movement, *Electroenceph clin Neurophysiol*, 103, pp. 642-651.
- Pfurtscheller, G. & Aranibar, A. (1977). Event related cortical desynchronization detected by power measurements of scalp EEG. *Electroenceph Clin Neurophysiol*, Vol. 42, pp. 817-826.
- Pfurtscheller, G.; Neuper, C.; Schlogl, A. & Lugger, K. (1998). Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Trans. on Rehabilitation Engineering*, Vol. 6, pp. 316-325.
- Pfurtscheller, G.; Neuper C.; Ramoser H. & Müller-Gerking J. (1999a). Visually guided motor imagery activates sensorimotor areas in humans. *Neuroscience Letters*, Vol. 269, No. 3, pp. 153-156.
- Pfurtscheller, G. & Guger, C. (1999b). Brain-computer communication system: EEG-based control of hand orthosis in a tetraplegic patient. *Acta Chir. Austriaca*, Vol. 31 23-25.
- Pfurtscheller, G. (1999c). EEG Event-related Desynchronisation (ERD) and Event-related Synchronisation (ERS). In *Niedermeyer, pp. 958-967, E., Lopes da Silva, F.H. (Eds.), Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*, forth ed., Williams and Wilkins, Baltimore.

- Pfurtscheller, G.; Neuper, C.; Guger, C.; Harkam, W.; Ramoser, H.; Schlogl, A.; Obermaier, B. & Pgegenzer, M. (2000a). Current trends in Graz brain-computer interface (BCI) research. *IEEE Trans. on Rehabilitation Engineering*, Vol. 8, No. 2, pp. 216-219.
- Pfurtscheller, G.; Guger, C.; Müller, G.; Krausz, G. & Neuper, C. (2000b). Brain oscillations control hand orthosis in a tetraplegic, *Neurosci. Lett.*, Vol.292, pp. 211-214.
- Pfurtscheller, G. & Neuper, C. (2001). Motor imagery and direct brain-computer communication. *proceedings of the IEEE*, Vol. 89, No. 7, pp. 1123-1134.
- Pfurtscheller, G.; Müller, G. R. ; Pfurtscheller, J.; Gerner, H. J. & Rupp, R. (2003). 'Thought' - control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neuroscience Letters*, V. 351, Issue 1, 6 November, pp. 33-36.
- Platt, J.; Cristanini, N. & Shawe-Taylor, J. (2000) . Large margin DAGs for multiclass classification. *Advances in Neural Information Processing Systems*, Vol. 12, pp. 543-557, MIT Press.
- Qin, L.; Ding, L. & He, B. (2004). Motor imagery classification by means of source analysis for brain computer interface applications. *Journal of Neural Engineering*, Vol. 1, No. 3, pp. 135-141.
- Qin, L. & He, B. (2005). A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications. *J. Neural Eng.*, Vol. 2, pp. 65-72.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, Vol. 77, pp. 257-286.
- Raudys, S. J. & Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 3, pp. 252-264,.
- Rakotomamonjy, A.; Guigue, V.; Mallet, G. & Alvarado, V. (2005). Ensemble of SVMs for improving brain computer interface p300 speller performances. *Int. Conf. on Artificial Neural Networks*.
- Ramoser, H.; Müller-Gerking, J. & Pfurtscheller, G. (2000). Optimal spatial filtering of singletrial EEG during imagined hand movement. *IEEE Trans. Rehab. Eng.*, Vol. 8, pp. 441-446.
- Roberts, S. J. & Penny, W. D. (2000). Real-time brain-computer interfacing: a preliminary study using Bayesian learning. *Med Biol Eng Comput.*, Jan, Vol. 38, No. 1, pp. 56-61.
- Rubinstein, Y. D. & Hastie, T. (1997). Discriminative versus informative learning. In: *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, pp. 49-53.
- Salimi Khorshidi, G., Jaafari, A., Motie Nasrabadi, A., Hashemi Golpayegani, M. (2007). Modifying the Classic Peak Picking Technique Using a Fuzzy Multi Agent to Have an Accurate P300-based BCI. *EUSFLAT Conf. (2)* pp. 143-147.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, Vol. 6, No. 2, pp. 461-464.
- Schlögl, A.; Flotzinger, D. & Pfurtscheller, G. (1997). Adaptive autoregressive modelling used for single-trial EEG classification. *Biomed. Technik*, Vol. 42, pp. 162-167.

- Scherer, R.; Müller, G. R.; Neuper, C.; Graimann, B. & Pfurtscheller, G. (2004). An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. *IEEE Trans. Biomed. Eng.*, Vol. 51, pp. 979-984.
- Schlögl, A.; Lee, F.; Bischof, H. & Pfurtscheller, G. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition. *J. Neural Eng.*, Vol. 2, pp. L14-L22.
- Scholkopf, B.; Mika, S.; Burges, C. J. C.; Knirsch, P.; müller, K. R.; Ratsch, G. & Smola, A. J. (1999). Input space versus feature space in kernel-based methods. *IEEE Trans. Neural Networks*, Vol. 10, No. 5, pp. 1000-1017.
- Schölkopf, B. & Smolan A. J. (2002). Learning with kernels. *The MIT Press*.
- Schlogl, Lugger, A.; K. & Pfurtscheller, G. (1997). Using adaptive autoregressive parameters for a brain-computer-interface experiment. In: *Proceedings 19th International Conference IEEE/EMBS*, pp. 1533-1535.
- Solhjoo, S., Nasrabadi, A. M. & Golpayegani, M. R. H. (2005). Classification of chaotic signals using HMM classifiers: EEG-based mental task classification. *Proc. European Signal Processing Conference*.
- Stanny, R. R. (1989). Mapping the event related Potentials of the brain: Theoretical issues, Technical considerations and computer programs. *Naval aerospace Medical Research Laboratory, Naval Air Station, Pensacola, Florida 32506-5700*.
- Sun, S.; Lan, M. & Lu, Y. (2008). Adaptive EEG signal classification using stochastic approximation methods. *Proceedings of the 33rd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 413-416.
- Thomas C. (1977). Firefox, New York, NY: Holt, Rinehart and Winston.
- Trad, D.; Helmy, S.; Al-ani, T.; Delaplace, S. (2009). Hidden Markov Models and Support Vector Machines applied to P300 Based Brain-Computer Interface - a Comparative study. *ICTA 09, Hammamet, 7-9 Mai*.
- Urszula, S. T.; Urszula, M-k & Kozik, A. (1999). Blinking Artefact Recognition in EEG Signal Using Artificial Neural Network. In: *Fourth Conference Neural networks and Their Applications, Czestochowa*.
- Vapnik, V. N. (1995). The nature off Statistical Learning Theory. *Springer-Verlag*.
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Trans. Neural Netw.*, Vol. 10, pp. 988-999.
- Varsta, M.; Heikkonen, J.; Millan, J. D. R. & Mourino, J. (2000). Evaluating the performance of three feature sets for brain-computer interfaces with an early stopping MLP commitee. *Int. Conf. on Pattern Recognition, (Barcelona, Spain)*, pp. 907-910.
- Vidal, J. J. (1973), Toward direct brain-computer communication. In: *Annual Review of Biophysics and Bioengineering*, L. J. Mullins (Ed.), Annual Reviews, Inc., Palo Alto, Vol. 2, pp. 157-80.
- Vidal, J. J. (1977), Real-time detection of brain events in EEG. In: *Proceedings of the IEEE*, May 1977, Vol. 65, No. 5, pp.633-641.
- Vidaurre, C.; Schlögl, A.; Cabeza, R.; Scherer, R. & Pfurtscheller G. (2007). Study of On-Line Adaptive Discriminant Analysis for EEG-Based Brain Computer Interfaces. *IEEE Transactions On Biomedical Engineering*, Vol. 54, No. 3, March, pp. 550-556.

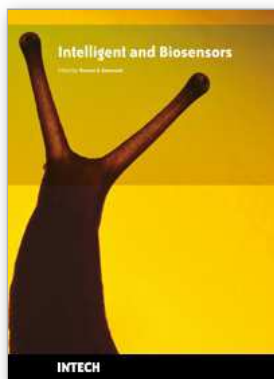
- Vourkas, M.; Papadourakis, G. & Micheloyannis, S. (2000). Use of ann and hjorth parameters in mental-task discrimination. In: *First International Conference on Advances in Medical Signal and Information Processing*, pp. 327-332.
- Wang, Y.; Berg, P. & Scherg, M. (1999). Common spatial subspace decomposition applied to analysis of brain responses under multiple task conditions: a simulation study. *IEEE Trans. Biomed. Eng.*, Vol. 110, Issue 4, pp. 604-614.
- Wang, T.; Deng, J. & He, B. (2004a). Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns. *Clin. Neurophysiol.*, Vol. 115 2744-2753.
- Wang, T.; Deng, J. & He, B. (2004b). Classification of motor imagery EEG patterns and their topographic representation. *Proc. 26th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, (San Francisco, CA), pp. 4359-4362.
- Weber, D.L. (2001) Scalp current density and source current modelling. http://dnl.ucsf.edu/users/dweber/dweber_docs/ EEG_scd.html. Visited on September, 10, 2009.
- Wessberg, J.; Stambaugh, C. R.; Kralik, J. D.; Beck, P. D.; Laubach, M.; Chapin, J. K.; Kim, J.; Biggs, S. J.; Srinivasan, M. A. & Nicolelis, M. A. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates, *Nature*, 16 Nov, Vol. 408, pp. 361-365.
- Wentrup, M. G.; Gramann, K.; Wascher, E. & Buss, M. (2005). Eeg source localization for brain-computer-interfaces. In: *2nd International IEEE EMBS Conference on Neural Engineering*, pp. 128-131.
- Wolpaw, J. R.; McFarland, D. J. & Vaughan, T. M. (2000). Brain-computer interfaces for communication and control, *IEEE Trans. on Rehabilitation Engineering*, June, Vol. 8, No. 2, pp. 222-226.
- Wolpaw, J. R.; Birbaumer, N.; McFarland, D. J.; Pfurtscheller, G. & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control, *Clin. Neurophysiol*, 113, pp. 767-91.
- Xu, B-G. & Song, A-G. (2008). Pattern recognition of motor imagery EEG using wavelet transform. *J. Biomedical Science and Engineering*, Vol. 1, pp. 64-67.

11. Appendix

Subj./Task	Task1	Task2	Task3	Task4	Task5	Task6
1/1	50	0	10	10	20	10
1/2	10	40	10	30	10	0
1/3	0	10	50	0	20	20
1/4	0	0	10	60	20	10
1/5	0	0	10	30	60	0
1/6	0	0	10	30	20	40
2/1	60	20	10	10	0	0
2/2	30	50	10	10	0	0
2/3	0	0	50	30	20	0
2/4	30	10	0	60	0	0
2/5	30	20	10	0	40	0
2/6	30	10	0	0	10	50
3/1	40	10	10	20	20	0
3/2	0	30	20	20	20	10
3/3	0	0	60	20	10	10
3/4	0	0	20	60	20	0
3/5	0	10	10	20	50	10
3/6	0	0	10	10	30	50
4/1	30	20.00	10	10	10	20
4/2	20	60	0	0	0	20
4/3	0	30	40	0	10	20
4/4	0	30	0	40	20	10
4/5	0	20	0	0	60	20
4/6	20	20	0	10	0	50
5/1 - 5/6	-	-	-	-	-	-
6/1	40	0	30	20	10	0
6/2	0	40	10	10	10	30
6/3	20	0	70	10	0	0
6/4	10	0	20	60	10	0
6/5	0	0	10	20	20	20
6/6	0	0	20	20	10	50
7/1	40	0	0	20	10	30
7/2	0	50	0	30	20	0
7/3	0	0	70	30	0	0
7/4	0	40	10	50	0	0
7/5	0	0	10	20	50	20
7/6	10	0	0	30	10	50
8/1	60	10	0	0	30	0
8/2	20	40	0	10	20	10
8/3	20	0	50	0	30	0
8/4	20	0	0	40	30	10
8/5	20	10	10	0	50	10
8/6	10	10	0	0	30	50
9/1	40	10	20	0	20	10
9/2	0	30	20	10	20	20
9/3	0	10	50	0	10	30
9/4	0	0	20	40	30	10
9/5	0	10	30	10	40	10
9/6	10	10	20	0	10	50

Task1	Task2	Task3	Task4	Task5	Task6
66	6	6	6	3	13
21	47	8	6	3	15
128	3	49	6	3	11
23	8	7	45	3	14
31	4	4	6	43	12
28	3	9	6	2	52
58	7	8	8	10	9
9	48	11	8	14	10
9	9	55	8	11	8
6	5	6	72	6	5
7	6	7	6	67	7
8	7	7	5	8	65
53	10	6	15	8	8
11	50	8	15	7	9
11	11	53	10	7	8
13	10	10	55	5	7
13	7	8	15	48	9
21	6	5	15	6	47
51	9	6	7	12	15
8	51	10	8	12	11
12	12	52	7	9	8
7	8	7	48	10	20
10	7	8	7	55	13
5	8	10	10	13	53
-	-	-	-	-	-
45	27	10	7	9	2
11	48	12	11	11	7
10	4	55	10	13	8
10	8	13	49	11	9
68	13	19	7	48	5
11	4	7	10	21	47
49	10	9	10	10	12
7	51	11	11	7	13
9	12	51	10	8	10
11	11	10	48	10	10
9	12	10	9	49	11
11	8	9	8	12	52
54	9	11	7	9	10
14	47	12	9	9	9
10	4	49	13	12	12
13	4	17	49	10	7
13	7	10	9	50	11
12	8	13	8	13	46
51	10	6	9	15	9
12	48	9	10	11	10
9	8	52	13	12	6
8	11	10	49	16	6
9	12	9	9	51	10
10	8	8	10	16	48

Table 1. Confusion matrices for the eight subjects and the six tasks (subject 5 is not taken into account). Left: *HMMs*-based, right: *M-SVM*-based. All the values are rounded to make the text readable.



Intelligent and Biosensors

Edited by Vernon S. Somerset

ISBN 978-953-7619-58-9

Hard cover, 386 pages

Publisher InTech

Published online 01, January, 2010

Published in print edition January, 2010

The use of intelligent sensors have revolutionized the way in which we gather data from the world around us, how we extract useful information from that data, and the manner in which we use the newly obtained information for various operations and decision making. This book is an attempt to highlight the current research in the field of Intelligent and Biosensors, thereby describing state-of-the-art techniques in the field and emerging new technologies, also showcasing some examples and applications.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Tarik Al-ani and Dalila Trad (2010). Signal Processing and Classification Approaches for Brain-Computer Interface, Intelligent and Biosensors, Vernon S. Somerset (Ed.), ISBN: 978-953-7619-58-9, InTech, Available from: <http://www.intechopen.com/books/intelligent-and-biosensors/signal-processing-and-classification-approaches-for-brain-computer-interface>

INTech
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821